

Gonçalo Valente da Silva Mariño Carrera

**Propriedades Físico-Químicas de
Líquidos Iónicos e estudos Químio-
Informáticos de Reactividade Química**

Dissertação apresentada para a obtenção do

Grau de Doutor em Química

Especialidade de Química Orgânica

Pela Universidade Nova de Lisboa

Faculdade de Ciências e Tecnologia

**Lisboa
2009**

- nº de arquivo

- “copyright”

Aos meus Pais

Agradecimentos

Aos meus pais pela constante presença na minha vida, por serem referências fundamentais nos valores e percurso de vida, pelo apoio dado de forma incondicional ao longo das suas vidas e em particular nesta fase. Aos meus pais devo, em grande medida, as minhas realizações, pelos preciosos ensinamentos que me foram dados.

Ao Professor Doutor João Aires de Sousa, orientador desta tese, pela constante e excelente orientação dada durante o decorrer do trabalho descrito nesta dissertação, pelo exemplo de elevado rigor científico demonstrado desde a primeira hora. Um agradecimento igualmente muito especial ao co-orientador desta tese o Professor Carlos Afonso, pelo exemplo de elevado rigor científico, pelos preciosos conselhos, demonstrando que os detalhes fazem toda a diferença.

Ao Doutor Luís Branco pelo acompanhamento dedicado ao trabalho realizado, pela constante cumplicidade demonstrada ao longo destes quatro anos, interessantes discussões e preciosos conselhos.

À Doutora Florbela Pereira minha colega de gabinete por todo o companheirismo demonstrado, interessantes discussões e preciosos conselhos.

À Doutora Maria Paz pelo companheirismo demonstrado e pelas interessantes “charlas” que tivemos.

Ao Doutor Sunil Gupta, Doutor Yuri Binev e Doutor Qingyou Zhang pelos preciosos conselhos dados.

Em geral ao grupo de trabalho do Professor Doutor Carlos Afonso pelo bom ambiente de trabalho criado, não querendo aqui particularizar nomes pois são realmente muitos podendo-me esquecer de referir algum.

Para o grupo de trabalho da Professora Doutora Ana Lobo, em particular o Doutor Ravi Varala, à Luísa Carvalho, Mónica Estevão e à Professora Doutora Ana Lourenço por nunca me terem negado o uso do laboratório, equipamento e material laboratorial na realização de alguns ensaios e preparação de amostras.

À Doutora Maria do Rosário Caras Altas pelo excepcional profissionalismo, simpatia e por ter colaborado sempre na obtenção de espectros de ^1H -RMN e ^{13}C -RMN em tempo útil.

Ao Doutor Júlio Santos e Doutora Teresa Maia pelo inestimável apoio dado, pelo excepcional profissionalismo, e pela acessibilidade a meios fundamentais na concretização deste objectivo.

À Doutora Cristina Fernandes e Doutora Janete Maximino pelo excepcional profissionalismo.

À minha família pelo apoio incondicional, em particular para os meus irmãos Marta Carrera e Rodrigo Carrera meus irmãos companheiros desde o primeiro momento. À minha avó Maria de Jesus e à minha tia Luzia da Silva por terem participado activamente na minha educação e por toda a amizade demonstrada desde a primeira hora.

Uma palavra especial para o meu avô Elias da Silva, pois apesar de já não se encontrar entre os vivos, foi sempre uma fonte inspiradora, um exemplo de integridade e rectidão moral que nunca esquecerei.

Aos meus avós paternos, Isolina Carrera Carrera e Ceferino Mariño Francisco, pois apesar de não os ter conhecido, muitas histórias inspiradoras ouvi a respeito dos seus percursos de vida e valores, transmitidos através do meu pai.

À Fundação para a Ciência e a Tecnologia pelo financiamento do trabalho presente nesta dissertação através da Bolsa de Doutoramento SFRH/BD/18354/2004.

Finalmente uma palavra de agradecimento a todos aqueles que de alguma forma contribuíram para a concretização deste objectivo.

Resumo

O trabalho descrito nesta dissertação compreende dois temas distintos. O primeiro consiste no estudo de propriedades físico-químicas de líquidos iónicos. O segundo é uma investigação quimio-informática de padrões de reactividade contidos em bases de dados de reacções orgânicas.

No primeiro tema foi estudado o ponto de fusão (p.f.) e densidade de líquidos iónicos. Um dos estudos consistiu na modelação do p.f./estado físico de sais de imidazólio sendo obtido modelo *Support Vector Machines* de classificação que permitiu a obtenção de razoável concordância (80%) para conjunto de teste. Características estruturais como o tamanho do catião, presença de ramificações, presença de átomos polarizáveis, cargas associadas a catião e anião, diferença de peso molecular entre catião e anião e tendência para formar pontes de hidrogénio emergiram dos modelos com árvores de decisão como factores relevantes para o p.f./estado físico. Foi também estudado o p.f. de sais de guanidínio sendo treinadas redes de *counterpropagation* que permitiram a obtenção de $R^2 = 0,815$ para o conjunto de teste. De acordo com indicações dos modelos construídos foram sintetizados novos sais de guanidínio e imidazólio cujos p.f. e/ou estados físicos correspondem às previsões obtidas pelos modelos. Noutro estudo foi modelada a densidade de sais de imidazólio, sendo treinado modelo baseado em *Support Vector Machines* que permitiu a obtenção de $R^2 = 0,975$ para o conjunto de teste. Nesse estudo foi ainda construída árvore de decisão que permitiu identificar como características estruturais relevantes, a presença de átomos pesados, a diferença de tamanho entre catião e anião, a presença de ramificações e átomos electronegativos. Foram ainda sintetizados novos sais de guanidínio na busca de sais com elevada densidade. Um dos compostos obtidos apresenta p.f. 74,3°C e densidade 1,68 g/cm³.

No segundo tema utilizaram-se descritores MOLMAP para a representação de compostos e reacções químicas. Estes descritores codificam os tipos de ligações covalentes presentes numa molécula e baseiam-se num mapa auto-organizativo que classifica ligações químicas de acordo com as suas propriedades físico-químicas e/ou topológicas. Com base num conjunto de reacções envolvendo butilamina foram construídos modelos de classificação de compostos e reacções segundo o mecanismo reaccional sendo obtidas concordâncias de 76% e 91% para os conjuntos de teste de compostos e reacções respectivamente. Foram ainda construídos dois modelos *Random Forest* para classificar compostos como reactivos ou não na presença de butilamina, num dos casos, e cianoborohidreto no segundo caso, sendo obtidas concordâncias de 88% e 94% para os respectivos conjuntos de teste.

Abstract

The work described in this dissertation comprises two main areas of research: a theoretical and experimental study of physicochemical properties of ionic liquids (IL), and the development of Chemoinformatics approaches in order to uncover reactivity patterns from databases of chemical reactions.

1) Study of melting point (mp) and density of ILs. One issue was the mp/physical state of imidazolium salts; a *Support Vector Machines*-based model was built and allowed a classification accuracy of 80% for the test set. The most relevant structural features for the mp/physical state, as revealed by decision tree, were the size and degree of branching of the cation, presence of polarizable atoms in the cation, maximum and minimum charges of both the cation and anion, difference of molecular weight between the cation and anion, as well as the potential of the cation to establish hydrogen bond interactions. In another study, quantitative structure-property relationships were established to predict the mp of guanidinium salts. Models were built with counterpropagation neural networks obtaining $R^2=0.815$ for the test set. Assisted by the models, new imidazolium and guanidinium salts were prepared, and the measured melting properties were in accordance with the predictions. Another physicochemical property studied was the density of imidazolium ionic liquids. Specifically a new *Support Vector Machines*-based model was obtained with excellent predictive power ($R^2=0.975$ - test set). Structural features selected as the most important by the decision trees models were the difference of sizes between cation and anion, branching, presence of electronegative and heavy atoms in the cation. Additionally new guanidinium salts of high density were synthesized. One of the synthesized compounds is an IL with mp of 74.3°C and density of 1.68 g/cm³.

2) Use of MOLMAP descriptors to encode compounds and chemical reactions. These descriptors encode the covalent bonds present in a molecule and are based in a self organizing map that classifies chemical bonds according to their physico-chemical and/or topological properties. MOLMAP descriptors were used to classify compounds and reactions according to reaction mechanism in the presence of butylamine; the selected models allowed accuracies of 76% and 91% for test sets considering compounds and reactions respectively. In another study two *Random Forest* models were built in order to classify compounds as reactive or not in the presence of butylamine, in one case, and sodium cyanoborohydride in another case. Concordances of 88% and 94% were obtained for the respective test sets.

Abreviaturas

AM1	- Austin Model 1
<i>b</i>	- Ordenada na origem
BOC	- <i>tert</i> -Butiloxicarbonil
BPG-NN	- Redes neuronais de Retropropagação (Back Propagation Neural Networks)
C	- Parâmetro de complexidade
CART	- Classification and Regression Trees
Cbz	- Carboxibenzil
<i>c.c.</i>	- Cromatografia em coluna simples
<i>c.c.f.</i>	- Cromatografia de camada fina
CPG-NN	- Redes neuronais de Counterpropagation
<i>c.v.</i>	- Validação cruzada
DTNB	- Ácido 5,5'-ditio-bis(2-nitrobenzóico)
EA	- Afinidade electrónica
ESCA	- Electron Spectroscopy for Chemical Analysis
FAB	- Fast Atom Bombardment
GAO	- Perfil das orbitais atómicas (Graphs of Atomic Orbitals)
GC	- Cromatografia Gasosa (Gás Chromatography)
HMO	- Teoria das orbitais moleculares de Hückel (Hückel Molecular Orbital theory)
HPLC	- Cromatografia Líquida de Alta eficiência (High Performance Liquid Chromatography)
<i>I(t)</i>	- Função de impureza
IP	- Potencial de ionização
IV	- Infra-Vermelho
<i>k_T</i>	- Compressibilidade isotérmica
LCAO	- Combinação Linear de Orbitais Atómicas (Linear Combination of Atomic Orbitals)
LCD	- Liquid crystal display
LI	- Líquido iónico
LLNA	- Local Lymph Node Assay
log <i>S_m</i>	- Logaritmo da solubilidade aquosa
Log <i>P</i>	- Logaritmo do coeficiente de partição entre octanol e água
<i>M</i>	- Massa molar

MAE	- Erro médio absoluto (Mean Absolute Error)
MALDI	- Desadsorção com laser assistida por matriz (Matrix Assisted Laser Desorption Ionization)
MLR	- Regressão multilinear (Multilinear Regression)
MOLMAP	- MOLEcular Maps of Atom-level Properties
MSE	- Erro médio quadrático
NEt₃	- Trietilamina
NHMe₂	- Dimetilamina
NTf₂⁻	- Anião bis(trifluorometilsulfonil)imida [(CF ₃ SO ₂) ₂ N ⁻]
OCDE	- Organização para a Cooperação e Desenvolvimento Económico
PCA	- Análise de componentes principais (Principal Component Analysis)
PEOE	- Partial Equalization of Orbital Electronegativity
<i>p.e.</i>	- Ponto de ebulição
PLS	- Mínimos quadrados parciais (Partial Least Squares)
PEPE	- Partial Equalization of π -electronegativity
<i>p.f.</i>	- Ponto de fusão
ppm	- partes por milhão
q_p	- Calor a pressão constante
q_{rev}	- Calor reversível
q_π	- Carga pi
q_σ	- Carga sigma
<i>R</i>	- Constante dos gases perfeitos
R^2	- Quadrado do coeficiente de correlação de Pearson
RAI	- Relative Alkylation Index
REACH	- Registration, evaluation, authorization of chemicals
RF	- <i>Random Forest</i>
R_I	- Índice de refração
R_M	- Refração molar
RMS	- Raiz do erro médio quadrático
$R_\alpha(T)$	- Medida de custo complexidade
SVM	- Máquinas de suporte vectorial (Support Vector Machines)
TSIL	- Líquido Iónico para Aplicação Específica (Task Specific Ionic Liquid)
<i>t.v.</i>	- Transição vítrea
<i>U</i>	- Energia Interna
<i>w</i>	- Trabalho realizado pelo sistema
α_i	- Multiplicador de Lagrange
α_p	- Coeficiente de expansão térmica

ΔS_{fus}	- Entropia de fusão
ΔH_{fus}	- Entalpia de fusão
ΔS_f^{conf}	- Contribuição da conformação para a entropia de fusão
ΔS_f^{trans}	- Contribuição translacional para a entropia de fusão
ΔS_f^{rot}	- Contribuição rotacional para a entropia de fusão
ΔG_{fus}^o	- Energia de Gibbs de fusão a condições de pressão e temperatura padrão
ΔG_{fus}^T	- Energia de Gibbs de fusão a uma temperatura genérica T
ΔG_{fus}	- Energia de Gibbs de fusão à temperatura de fusão
ϵ	- Constante dielétrica
ϵ	- Desvio autorizado em algoritmo Support Vector Machines de Regressão
η_i	- Multiplicador de Lagrange
ξ_i	- Desvio adicional
ρ	- Densidade
σ	- Simetria Rotacional
τ	- Flexibilidade molecular
χ_σ	- Electronegatividade sigma
χ_π	- Electronegatividade pi
$^{13}\text{C-RMN}$	- Ressonância magnética nuclear de carbono
$^1\text{H-RMN}$	- Ressonância magnética nuclear de protão

Índice de Matérias

Agradecimentos	VII
Resumo	IX
Abstract	XI
Abreviaturas	XIII

Capítulo I

Métodos de aprendizagem automática e descritores moleculares	1
I-1: MÉTODOS DE APRENDIZAGEM AUTOMÁTICA	3
I-1.1: Redes Neurais	3
I-1.1.1: Redes Neurais de <i>Counter-Propagation</i>	6
I-1.1.2: Redes Neurais de <i>Kohonen</i>	10
I-1.2: Árvores de Decisão	12
I-1.3: <i>Random Forest</i>	15
I-1.4: <i>Support Vector Machines</i>	18
I-1.4.1: <i>Support Vector Machines</i> para Classificação.....	19
I-1.4.2: <i>Support Vector Machines</i> para Regressão	25
I-2: DESCRITORES.....	29
I-2.1: Descritores Calculados pelo Programa Dragon.....	30
I-2.2: Descritores Baseados no Programa PETRA	35
I-2.2.1: PETRA	36
I-2.2.2: Função de Distribuição Radial - RDF	51
I-2.2.3: Descritores SURFACE.....	51
I-2.2.4: Descritores XPET	53
I-2.3: Descritores Chemaxon	53
I-3: MÉTODOS DE SELECÇÃO DE DESCRITORES.....	53
Algoritmos Genéticos.....	54
I-4: BIBLIOGRAFIA.....	56

Capítulo II

Estudo de propriedades físico-químicas de líquidos iónicos.....	59
II-1: INTRODUÇÃO TEÓRICA.....	61
II-1.1: Enquadramento.....	61
II-1.2: Líquidos Iónicos	62
II-1.2.1: Aspectos Gerais de Líquidos Iónicos.....	62
II-1.2.1.1: Líquidos Iónicos em Síntese Orgânica.....	64
II-1.2.1.2: Líquidos Iónicos como Cristais Líquidos e Outros Materiais	65
II-1.2.1.3: Líquidos Iónicos em Extracções	68
II-1.2.1.4: Líquidos Iónicos Noutras Aplicações.....	69
II-1.2.1.5: Líquidos Iónicos Implementados na Indústria	70
II-1.2.2: Sais de Imidazólio	70
II-1.2.3: Sais de Guanidínio	74
II-1.2.4: Obtenção de Líquidos Iónicos Purificados	82
II-1.3: Propriedades Físico-Químicas de Líquidos Iónicos	83
II-1.3.1: Ponto de Fusão	83
II-1.3.1.1: Ponto de Fusão de Líquidos Iónicos.....	89

II-1.3.2: Densidade	97
II-1.3.2.1: Densidade de Líquidos Iônicos: Perspectiva Sintética	100
II-1.3.2.2: Densidade de Líquidos Iônicos: Métodos de Previsão	103
II-2: METODOLOGIA COMPUTACIONAL	105
II-2.1: Conjuntos de Dados	105
II-2.1.1: Representação das Estruturas Químicas por Descritores Moleculares	105
II-2.1.2: Normalização de Dados	119
II-2.1.3: Representação de sais	119
II-2.1.4: Características Específicas dos conjuntos de dados	120
II-2.2: Modelos Construídos	122
II-2.2.1: Estudo A - Modelação de Pontos de Fusão de Sais de Imidazólio.....	122
II-2.2.1.1: Modelo <i>Random Forest</i>	123
II-2.2.1.2: Modelos <i>Support Vector Machines</i>	123
II-2.2.1.3: Modelo Árvore de Decisão	123
II-2.2.2: Estudo B - Modelos para Classificação de Sais de Imidazólio segundo Estado Físico/Ponto de Fusão	123
II-2.2.2.1: Modelo <i>Random Forest</i>	124
II-2.2.2.2: Modelos <i>Support Vector Machines</i>	124
II-2.2.2.3: Modelo Árvore de Decisão	124
II-2.2.3: Estudo C - Modelação da Densidade de Sais de Imidazólio.....	124
II-2.2.3.1: Modelo <i>Random Forest</i>	125
II-2.2.3.2: Modelos <i>Support Vector Machines</i>	125
II-2.2.3.3: Modelo Árvore de Decisão	125
II-2.2.4: Estudo D - Modelação do Ponto de Fusão de Sais de Guanidínio	125
II-2.2.4.1: Modelo Baseado numa Rede de <i>Counter-Propagation</i>	126
II-2.2.4.2: Modelo baseado num Conjunto de 5 redes de <i>Counter-Propagation</i>	126
II-2.2.4.3: Modelo baseado em Rede de <i>Counter-Propagation</i> Treinada com Conjunto Seleccionado de Descritores.....	127
II-2.3: Cálculo do Valor de Densidade e Ponto de Fusão Previsto por Modelos Computacionais	127
II-2.4: Medidas de Qualidade dos Modelos de Previsão	128
II-2.4: Estudo A - Modelação de Pontos de Fusão de Sais de Imidazólio.....	128
II-2.4: Estudo B - Modelos para Classificação de Sais de Imidazólio segundo estado Físico/Ponto de Fusão	129
II-2.4: Estudo C - Modelação da Densidade de Sais de Imidazólio.....	130
II-2.4: Estudo D - Modelação do Ponto de Fusão de Sais de Guanidínio	130
II-3: RESULTADOS E DISCUSSÃO	131
II-3.1: Estudo A - Modelação do Ponto de Fusão de Sais de Imidazólio	131
II-3.2: Estudo B - Classificação de Sais de Imidazólio segundo Ponto de Fusão/Estado físico	138
II-3.3: Estudo C - Modelação da Densidade de Sais de Imidazólio.....	155
II-3.4: Estudo D - Modelação do Ponto de Fusão de Sais de Guanidínio	164
II-3.5: Estudo E - Síntese e Estudo de Sais de Guanidínio de Elevada Densidade	181
II-3.6: Estudo F - Síntese de sais de Guanidínio contendo Tiocianato, <i>para</i> -Toluenosulfonato ou Diclanamida como Aniões e Estudo alargado da Influência de Cátion e Anião no Ponto de Fusão	213
II-4: CONCLUSÃO.....	228

II-5: PARTE EXPERIMENTAL.....	231
II-5.1: Preâmbulo.....	231
II-5.2: Síntese de Sais de Imidazólio indicados por Modelo de Previsão de Pontos de Fusão.....	234
II-5.2.1: Brometo de 1-benzil-3-propil imidazólio (1a)	234
II-5.2.2.1: Tetrafluoroborato de 1-benzil-3-propil imidazólio (2a).....	234
II-5.2.2: Compostos Resultantes da troca Aniônica Realizada Partindo de Brometo de 1-benzil-3-propil imidazólio (1a)	234
II-5.2.2.2: Trifluorometanosulfonato de 1-benzil-3-propil imidazólio (3a)....	235
II-5.3: Síntese de Sais de Guanidínio Indicados por Modelo de Previsão de Pontos de Fusão.....	236
II-5.3.1: Síntese de cloretos de Guanidínio.....	236
II-5.3.1.1: Método Geral	236
II-5.3.1.1.1: Cloreto de <i>N,N,N',N''</i> -tetrametil- <i>N',N''</i> -dioctadecilguanidínio (4a)	236
II-5.3.1.1.2: Cloreto de <i>N,N</i> -dimetil- <i>N',N''</i> -di-hexil- <i>N',N''</i> -bis(2-etilhexil)guanidínio (5a)	237
II-5.3.1.1.3: (2-Etil-hexil)-hexilamina (5b)	238
II-5.3.1.1.4: Cloreto de <i>N,N</i> -dimetil- <i>N',N'',N''</i> -tetraalilguanidínio (6a)... ..	238
II-5.3.2: Síntese de Sais de tetrafenilborato de Guanidínio	239
II-5.3.2.1: Método Geral	239
II-5.3.2.1.1: Tetrafenilborato de <i>N, N, N',N''</i> -Tetrametil- <i>N',N''</i> -dioctadecilguanidínio (7a).....	239
II-5.3.2.1.2: Tetrafenilborato de <i>N,N</i> -dimetil- <i>N',N'',N'',N''</i> -tetrahexilguanidínio (8a).....	240
II-5.3.2.1.3: Cloreto de <i>N,N</i> -dimetil- <i>N',N'',N'',N''</i> -tetrahexilguanidínio (8b)	241
II-5.3.2.1.4: Tetrafenilborato de <i>N,N</i> -Dimetil- <i>N',N''</i> -di-hexil- <i>N',N''</i> -bis(2-etilhexil)guanidínio (9a)	241
II-5.4: Síntese de Líquidos Iônicos Densos, Sais de Guanidínio:	242
II-5.4.1: Compostos Derivados do cloreto de <i>N,N</i> -dimetil- <i>N',N'',N''</i> -tetraalilguanidínio (6a).....	242
II-5.4.1.1: Hexafluorofosfato de <i>N,N</i> -dimetil- <i>N',N'',N'',N''</i> -tetraalilguanidínio (10a)	242
II-5.4.1.2: Hexafluorofosfato de <i>N,N</i> -dimetil- <i>N',N'',N'',N''</i> -tetraquis(2,3-dibromopropil) guanidínio (11a).....	243
II-5.4.1.3: Bis(trifluorometanosulfonil)imidato de <i>N,N</i> -dimetil- <i>N',N'',N'',N''</i> -tetraalilguanidínio (12a).....	244
II-5.4.1.4: Bis(trifluorometanosulfonil)imidato de <i>N,N</i> -dimetil- <i>N',N'',N'',N''</i> -tetraquis(2,3-dibromopropil) guanidínio (13a).....	245
II-5.4.1.5: Perfluorooctanoato de <i>N,N</i> -dimetil- <i>N',N'',N'',N''</i> -tetraalilguanidínio (14a).246	
II-5.4.2: Compostos Derivados do cloreto de <i>N,N</i> -dimetil- <i>N',N''</i> dietil - <i>N',N''</i> -bis(2-metilalil)guanidínio (15a)	247
II-5.4.2.1: Cloreto de <i>N,N</i> -dimetil- <i>N',N''</i> dietil , <i>N',N''</i> -bis(2-metilalil)guanidínio (15a)	247
II-5.4.2.2: 1-etil-3,3-dimetil-1-(2-metilalil)-ureia (15b)	248
II-5.4.2.3: Bis(trifluorometanosulfonil)imidato de <i>N,N</i> -dimetil- <i>N',N''</i> dietil - <i>N',N''</i> -bis(2-metilalil)guanidínio (16a)	248
II-5.4.2.4: Bis(trifluorometanosulfonil)imidato de <i>N,N</i> -dimetil- <i>N',N''</i> dietil - <i>N',N''</i> -bis(2,3-dibromo-2-metil-propil) guanidínio (17a).....	249
II-5.4.3: Síntese do cloreto de <i>N,N</i> - dimetil – <i>N',N''</i> -dialil- <i>N',N''</i> diocetilguanidínio (18a).....	250

II-5.4.3.1: Cloreto de <i>N,N</i> - dimetil – <i>N',N''</i> -dialil- <i>N',N''</i> -dioctilguanidínio (18a)	251
II-5.4.3.2: 1-alil-3,3-dimetil-1-octilureia (18b)	251
II-5.4.3.3: Alil-octilamina (18c)	252
II-5.5: Síntese de Sais de Guanidínios: Baseados no Anião Dicianamida, Tiocianato e para-Toluenosulfonato	252
II-5.5.1.1: Tiocianato de <i>N, N, N',N''</i> -Tetrametil- <i>N',N''</i> -dioctadecilguanidínio (19a)	253
II-5.5.1: Síntese de Sais Baseados no Catião de <i>N, N, N',N''</i> -Tetrametil- <i>N',N''</i> -dioctadecilguanidínio	253
II-5.5.1.2: Dicianamida de <i>N, N, N',N''</i> -Tetrametil- <i>N',N''</i> -dioctadecilguanidínio (20a)	253
II-5.5.2: Síntese de Sais Baseados no Catião <i>N,N</i> -dimetil- <i>N',N'',N'''</i> -tetraalilguanidínio	254
II-5.5.2.1: Dicianamida de <i>N,N</i> -dimetil- <i>N',N'',N'''</i> -tetraalilguanidínio (21a)	254
II-5.5.2.2: Tiocianato de <i>N,N</i> -dimetil- <i>N',N'',N'''</i> -tetraalilguanidínio (22a)	255
II-5.5.3: para-Toluenosulfonato de <i>N,N</i> -dimetil- <i>N',N'',N'''</i> -tetrahexilguanidínio (23a)	255
II-6 – ANEXOS:	257
II-7 - BIBLIOGRAFIA:	300
Capítulo III	
Aplicação de Métodos de Aprendizagem Automática e Descritores MOLMAP a Problemas de Reactividade Química	305
III-1: INTRODUÇÃO TEÓRICA	307
III-1.1: Enquadramento	307
III-1.2: Classificação de Reacções Químicas	308
III-1.3: Estado da Arte sobre Métodos Químio-Informáticos para a Previsão de Reactividade Química	309
III-1.3.1: Previsão de Reactividade em Reacções Orgânicas	309
III-1.3.2: Previsão de Reactividade em Reacções Metabólicas	311
III-1.3.3: Previsão de Toxicidade e outras Actividades Biológicas	312
III-1.4: Descritores MOLMAP	323
III-1.5: Derivação de Informação Negativa a Partir de Bases de Dados Reaccionais Compostas Exclusivamente por Reacções que Ocorrem	326
III-2: METODOLOGIA COMPUTACIONAL	327
III-2.1: Bases de Dados Reaccionais Originais	327
III-2.1.1: Estandardização de Reacções	327
III-2.1.2: Selecção de Reacções	328
III-2.2: Descritores Moleculares MOLMAPs	329
III-2.2.1: Mapeamento de ligações	329
III-2.2.2: Obtenção de Descritores MOLMAPs	339
III-2.3: Conjuntos de Dados Derivados de Bases de Dados Reaccionais	342
III-2.4: Modelos Construídos	348
III-2.4.1. a): Classificação de Compostos de Acordo com o Mecanismo de Reacção com a butilamina	348
III-2.4.2 b): Classificação de reacções com butilamina	351
III-2.4.3 c): Estudo de Reactividade de Compostos na Presença de butilamina	351

III-2.4.4 d): Estudo de Reactividade de Compostos na Presença de Cianoborohidreto de Sódio	354
III-2.5: Medidas de Qualidade dos Modelos de Previsão	356
III-3: RESULTADOS E DISCUSSÃO	357
III-3.1 a): Classificação de Compostos quanto ao Mecanismo da Reacção com butilamina	357
III-3.2 b): Estudo do Mecanismo Reaccional de Reacções na Presença de butilamina	383
III-3.3 c): Estudo de Reactividade de Compostos na Presença de butilamina	394
III-3.4 d): Estudo de Reactividade de Compostos na Presença de cianoborohidreto de sódio	421
III-4: CONCLUSÃO	432
 Capítulo IV 	
Glossário.....	437

Índice de Figuras

Capítulo I

Figura I.1: Estrutura de um neurónio e interacção com outro neurónio através de sinapses.	3
Figura I.2: Representação da incorporação e processamento de um determinado sinal de <i>input</i> por parte de uma rede neuronal <i>Backpropagation</i>	5
Figura I.3: Treino e previsão numa rede neuronal <i>Counter-Propagation</i> que aprende relações entre descritores moleculares (camada de <i>input</i>) e propriedade a modelar (camada de <i>output</i>) durante a fase de treino. Na fase de previsão, vão ser obtidas previsões para novos objectos.	7
Figura I.4: Perfis mais comuns da função $a(dc-di)$, com valores máximos para o neurónio vencedor. Três tipos de funções diferentes: a) constante, b) triangular, c) chapéu mexicano	9
Figura I.5: Processo de treino de uma rede de <i>Kohonen</i> em que um objecto n dimensional está a ser submetido à rede, levando à correcção de uma série de pesos do neurónio vencedor e neurónios vizinhos.	11
Figura I.6: A aparente forma quadrangular (rectangular) das redes neuronais de <i>Kohonen</i> e de <i>Counter-Propagation</i> representa na realidade um topologia toroidal.	12
Figura I.7: Representação esquemática de uma árvore de decisão na resolução de um problema de regressão	13
Figura I.8: Conjunto de árvores de decisão, constituindo uma <i>Random Forest</i>	16
Figura I.9: Exemplo de duas classes linearmente separáveis	20
Figura I.10: Caso de problema não separável linearmente no espaço de <i>input</i> original a) mas linearmente separável num novo espaço de maior dimensionalidade onde os objectos são projectados segundo $\Phi(x)$ b)	22
Figura I.11: Variação do grau de complexidade da fronteira entre dois tipos de objectos, desde o caso mais complexo representado em a) até ao caso mais simples d)	24
Figura I.12: Exemplo de função linear onde desvios superiores a ε são tolerados a) sendo ζ o excesso em relação a ε . À direita b) aparece um tipo de função penalizadora que lida com desvios superiores a ε	26
Figura I.13: Tipos de efeitos contabilizados pela componente topológica do peso que determinada estrutura de ressonância vai ter na obtenção de cargas σ . a) alteração no número de ligações covalentes, b) alterações em sistemas aromáticos e c) contabilização da separação de cargas.	38
Figura I.14: As duas direcções consideradas na quebra heterolítica de uma ligação química.	46
Figura I.15: Quebra heterolítica de ligação b e representação da deslocalização de carga em fragmentos gerados.	48
Figura I.16: Estabilização de carga por parte de um átomo específico (neste caso N) influenciada por átomos vizinhos em função da sua proximidade ao átomo em causa. Essa medida de estabilização denomina-se polarizabilidade efectiva.	50
Figura I.17: Obtenção da superfície molecular tendo em conta o volume de Van der Waals ocupado por cada átomo da molécula e uma esfera de solvente de raio r	52
Figura I.18: Superfície molecular definida por uma determinada densidade de pontos.	52
Figura I.19: Determinado cromossoma representa um sub-conjunto de descritores. O “0” significa ausência de descritor e “1” significa presença de descritor.	54
Figura I.20: A) Crossover entre cromossomas parentais, para obtenção dos cromossomas descendentes. B) Mutação de um gene.	55

Capítulo II

Figura II.1: Estruturas catiónicas habituais em líquidos iónicos.....	62
Figura II.2: Estruturas aniónicas habituais em líquidos iónicos.....	63
Figura II.3: Estrutura genérica de um alcanosulfonato de guanidínio.	67
Figura II.4: Estrutura genérica de alquilbenzosulfonatos 17 e alquilbifenilsulfonatos 18 de guanidínio.	67
Figura II.5: As quatro possíveis estruturas de ressonância de um catião guanidínio.	75
Figura II.6: Estrutura do aminoácido arginina.	75
Figura II.7: Representação da peculiar interacção entre unidade guanidínio e carboxilatos.	76
Figura II.8: Isomeria E/Z num sal de guanidínio.	81
Figura II.9: O efeito que pequenas subtilezas estruturais podem ter no ponto de fusão de sais de piridínio.	91
Figura II.10: Estrutura genérica de um benzoimidazólio.....	92
Figura II.11: Ciclo de Born-Fajans-Haber para determinar ΔG_f^T	94
Figura II.12: Ciclo Born-Fajans modificado para cálculo de ΔG_f^T	96
Figura II.13: Sal de guanidínio vs sal de imidazólio.....	102
Figura II.14: Estrutura de um catião guanidínio evidenciando os seus seis grupos substituintes.	115
Figura II.15: Duas estruturas de ressonância de um catião imidazólio com a carga positiva a estar centrada nos dois átomos de azoto.....	119
Figura II.16: As três estruturas de ressonância mais relevantes num catião baseado na unidade guanidínio com a carga positiva a estar centrada nos átomos de azoto.	120
Figura II.17: Conjunto de teste de 79 sais – ponto de fusão experimental vs ponto de fusão previsto por modelo RF global e respectivo quadrado do coeficiente de correlação de Pearson R^2	138
Figura II.18: Distribuição de sais por anião presente e classe ponto de fusão/estado físico. BF4 – tetrafluoroborato, Br – brometo, Cl – cloreto, BTfI – bis(trifluorometanosulfonil)imidato, PF6 – hexafluorofosfato e CF3SO3 trifluormetanosulfonato.	142
Figura II.19: Catião de imidazólio e os seus 5 grupos substituintes.	143
Figura II.20: Árvore de Decisão construída para previsão de classe ponto de fusão/estado físico de sais de imidazólio.	144
Figura II.21: Relação entre o patamar de probabilidade mínima para considerar objectos e a concordância para o conjunto de teste.....	148
Figura II.22: Conjunto de teste - Relação entre o patamar mínimo para considerar objectos, relativamente ao valor da proximidade máxima com objectos do conjunto de treino, e a concordância.	149
Figura II.23: Duas estruturas alvo seleccionadas pelo modelos para serem líquidos iónicos.	149
Figura II.24: Figura esquemática do modelo Árvore de Decisão usado na modelação da densidade de sais de imidazólio.	158
Figura II.25: Conjunto de teste – correlação entre densidades experimentais e densidades previstas por modelo SVM.....	163
Figura II.26: Comparação entre pontos de fusão experimentais e pontos de fusão previstos por modelo baseado em <i>ensemble</i> de 5 redes CPG – Conjunto de treino e conjunto de teste. Em ambos os casos é determinado o quadrado do coeficiente de correlação de Pearson R^2	166
Figura II.27: Estruturas de sais de guanidínio mapeadas em modelo baseado numa rede de <i>Counter-Propagation</i>	169
Figura II.28: Representação e comparação das quatro camadas de <i>output</i> de modelo baseado numa rede neuronal de <i>Counter-Propagation</i> , correspondendo a	

	quatro famílias aniónicas, (a) cloreto (Cl ⁻), (b) tetrafenilborato (BPh ₄ ⁻), (c) brometo (Br ⁻) e (d) iodeto (I ⁻). As cores codificam os pesos da camada de <i>output</i> – neurónios azuis representam pontos de fusão elevados, enquanto neurónios vermelhos representam pontos de fusão baixos...	170
Figura II.29:	Novos sais de guanidínio sintetizados, pontos de fusão previstos por modelo baseado em <i>ensemble</i> de cinco redes neuronais CPG e pontos de fusão experimentais (em parênteses).....	172
Figura II.30:	Comparação relativamente ao número de grupos substituintes contendo funcionalidades olefínicas passíveis de serem bromadas entre catiões guanidínio e imidazólio 1,3 disubstituído.	183
Figura II.31:	Espectro de massa (FAB+) de sólido isolado de reacções de bromação e troca de anião partindo de cloreto de <i>N,N</i> -dimetil- <i>N',N'',N''</i> -tetraalilguanidínio 6a. Aparece assinalada no topo do respectivo sinal a combinação maioritária de isótopos relativamente às estruturas apresentadas na Figura II.32.....	184
Figura II.32:	Estruturas catiónicas correspondentes aos sinais indicados na figura II.31.....	183
Figura II.33:	Espectro de massa (FAB+) de sólido isolado resultado de método de síntese modificado realizando sequencialmente reacção de troca de anião e posteriormente reacção de bromação. Aparece assinalada no topo do respectivo sinal a combinação maioritária de isótopos relativamente aos fragmentos obtidos a partir da estrutura 11a.	186
Figura II.34:	1-Alil-3,3-dimetil-1-octilureia – produto secundário 18b.....	205
Figura II.35:	Aniões usados na síntese de novos sais de guanidínios – 135 dicianamida, 136 <i>para</i> -toluenosulfonato e 137 tiocianato.	213
Figura II.36:	Cloretos de guanidínio sintetizados para o estudo d) e usados na síntese de novos sais de guanidínio contendo tiocianato, <i>para</i> -toluenosulfonato e dicianamida como aniões.	214
Figura II.37:	Trocas de anião para a obtenção dos sais baseados nos aniões tiocianato 19a e dicianamida 20a.	214
Figura II.38:	Reacções de troca de anião para a síntese dos sais de dicianamida 21a e tiocianato 22a.	217
Figura II.39:	Reacção de troca de anião para a síntese do sal contendo anião <i>para</i> -toluenosulfonato 23a.....	219
Figura II.40:	Micropicnómetro utilizado para medir densidade de líquidos em volumes inferiores a 700 µl.	232
Figura II.41:	Picnómetro usado para medir densidade de sólidos.....	233

Capítulo III

Figura III.1:	Representação esquemática das duas fases associadas ao aparecimento da dermatite de contacto alérgica: Sensibilização e Contacto posterior. ²² ...	315
Figura III.2:	Aminoácido L-lisina e o pormenor da sua cadeia lateral, semelhante à butilamina.....	322
Figura III.3:	Exemplo esquemático do processo de treino, que visa a obtenção de um mapa de ligações químicas como o da Figura III.4.....	324
Figura III.4:	Exemplo de um mapa de <i>Kohonen</i> após a fase do treino com uma série de ligações químicas mapeadas representadas pelos elementos químicos e pela ordem dessa ligação. Por exemplo a ligação CC_1 representa ligação simples carbono-carbono.	324
Figura III.5:	Perfil das ligações químicas de uma determinada molécula, representado cada tipo de ligação por um “x” no mapa de <i>Kohonen</i>	325

Figura III.6: Representação do perfil das ligações químicas de uma molécula na forma numérica em que cada número representa a abundância de determinado tipo de ligação.....	325
Figura III.7: 15 compostos contendo ligação carbonilo constituindo diferentes funcionalidades e abarcando diversos ambientes químicos.....	331
Figura III.8: Molécula tendo em evidência diferentes ligações carbono-carbono, ligações simples, duplas, triplas e aromáticas.....	331
Figura III.9: Molécula onde estão evidenciadas diversas ligações carbono-halógeno e carbono-oxigênio.....	332
Figura III.10: Estrutura onde estão evidenciados diferentes tipos de ligação carbono-azoto.....	332
Figura III.11: Molécula onde estão evidenciadas ligações carbono-azoto em aminas e grupos nitro, tanto em ambiente alifático como aromático.....	332
Figura III.12: Molécula apresentando diversas ligações carbono-azoto em aminas primárias, secundárias, terciárias e amidas.....	333
Figura III.13: Molécula apresentando diversos tipos de ligações carbono-oxigênio.....	333
Figura III.14: Composto aromático substituído.....	333
Figura III.15: Esquema de derivação de MOLMAP de pseudo-composto que não reage.....	345
Figura III.16: Funcionalidade oxi-metilamina.....	348
Figura III.17: Mapeamento de 15 carbonilos (C=O) em SOM 29×29 baseado em 59 propriedades físico-químicas/topológicas das ligações. Cor verde representa maioritariamente: ligações CC simples, duplas e aromáticas, CN simples e aromáticas, CP simples, CS simples, CH, CB simples, C-Sn simples e CS dupla. Cor azul pálida representa ligações NC simples, dupla, tripla e aromática, NO simples e NN dupla e aromática. Cor azul celeste: OC simples, dupla e aromática, OSi simples e OH simples. Cor azul: ligações ClC, BrC, IC e BC. Cor azul escura: ligações SC simples, dupla e aromática, SN simples, SP simples e dupla, SH simples e SS simples. Cor rosa: ligações FC.....	359
Figura III.18: Compostos contendo ligação carbonilo (C=O) mapeados na Figura III.17.....	360
Figura III.19: Árvore de Decisão construída com MOLMAPs 25x25 aplicada ao estudo de classificação de compostos quanto ao mecanismo de reacção na presença de butilamina.....	376
Figura III.20: Mapa de <i>Kohonen</i> onde estão mapeados conjunto de treino – lado esquerdo e conjunto de validação – lado direito. Rede de <i>Kohonen</i> treinada com 32 descritores MOLMAP de reagente seleccionados por algoritmos genéticos. Vermelho – classe A, Azul escuro – classe B, Verde – classe C, Amarelo – classe D, Azul claro – classe E e Rosa - classe F.....	382
Figura III.21: Reacções mal classificadas segundo modelo <i>Random Forest</i>	387
Figura III.22: Representação esquemática da Árvore de Decisão utilizada para classificar reacções químicas segundo mecanismo reaccional.....	389
Figura III.23: Representação esquemática de reacções de substituição nucleofílica e adição de Michael em que aparece evidente a formação de ligação azoto-carbono.....	390
Figura III.24: Exemplos de reacções de classe “f”, que ainda ficaram associadas ao nó 10.....	392

Figura III.25: Mapeamento de 15 carbonilos (C=O), referentes a estruturas representadas na Figura III.26 em SOM 29×29 baseado em 11 propriedades físico-químicas das ligações. Cor verde representa maioritariamente: ligações CC simples, duplas e triplas, CN simples, CSi simples, CH, ClC, BrC, CP simples, CS simples e ClS simples. Cor azul pálida representa ligações NC simples, dupla, tripla, NH, NO simples, NN simples e dupla e NP simples. Cor azul celeste: OC simples e dupla, OSi simples, OH simples, ON simples, OP simples e OS dupla. Cor azul: ligações ClC e BrC,. Cor azul escura: ligações SC simples, dupla, SN simples, SP dupla, SH simples. Cor rosa: ligações FC. Cor lilás: ligações HC.....	396
Figura III.26: 15 compostos contendo ligação carbonilo (C=O) constituindo diferentes funcionalidades e abrangendo diversos ambientes químicos.....	397
Figura III.27: Árvore de Decisão, relativa à classificação de compostos segundo a reactividade na presença de butilamina, construída com MOLMAPs 29×29 tipo 1_0 em que a geração de pseudocompostos que não reagem baseou-se na subtracção de MOLMAP da butilamina ao MOLMAP do Produto antes da comparação com MOLMAP de Reagente – classe “c” reage, classe “a” não reage. Assinalado a amarelo aparece descritor A835 associado a aminas, funcionalidade que não está associada à reactividade na presença de nucleófilos.	402
Figura III.28: Associação entre a reactividade experimental (c – reactivo, a – não reactivo) e a probabilidade de reacção designada pela RF para os 49 compostos do conjunto independente de teste. Previsões correctas assinaladas por barras a cinzento e previsões incorrectas assinaladas por barras brancas.	406
Figura III.29: Curva ROC obtida com os resultados do melhor modelo RF para os 49 compostos do conjunto independente de teste.	406
Figura III.30: Composto 141 que reage segundo mecanismo de substituição conjugada, semelhante ao mecanismo de adição conjugada representada pela adição de Michael.	410
Figura III.31: Estrutura usada na derivação de MOLMAP de pseudocomposto que não reage semelhante a MOLMAP de 112.	411
Figura III.32: Estrutura usada na obtenção de MOLMAP de pseudocomposto que não reage, semelhante a MOLMAP de 137.	412
Figura III.33: Concordância vs patamar mínimo de probabilidade de previsão para considerar objectos de conjunto de 32 compostos testados <i>in vitro</i> por Gerberick <i>et al.</i> ³	413
Figura III.34: Árvore de decisão, relativa à classificação de compostos segundo a reactividade na presença de butilamina, construída com MOLMAPs 20x20, em que a geração de pseudocompostos que não reagem baseou-se na comparação entre MOLMAP do Produto ao MOLMAP de Reagente de determinada reacção. Classe c – reage, classe a – não reage.	414
Figura III.35: Rede de Kohonen 20×20 treinada com selecção de 64 descritores MOLMAP (29×29 do tipo 10_1) onde foram mapeados os compostos pertencentes ao conjunto independente de teste. Os neurónios foram classificados de acordo com os objectos do treino que os activam: vermelho – não reage, verde – reage.....	419
Figura III.36: Tipos de reacção presentes na base de dados reaccional em que intervém o cianoborohidreto de sódio.....	422
Figura III.37: Curva ROC obtida com os resultados do melhor modelo RF para os 48 compostos do conjunto independente de teste.	425
Figura III.38: Esquema de árvore de decisão, construída com MOLMAPs 15x15 do tipo 1_0 usada para modelar a reactividade de composto na presença de cianoborohidreto de sódio. Classe c – reage, classe a – não reage.....	428

Figura III.39: Funcionalidades reactivas na presença de cianoborohidreto de sódio....430

Índice de Esquemas

Capítulo II

Esquema II.1: Alquilação do metilimidazol com um haleto de alquila, resultando num sal de imidazólio.....	71
Esquema II.2: Síntese de um sal de 1,3-dialquilimidazólio.....	71
Esquema II.3: Esquema da troca de anião de um haleto X por outro tipo de anião Y....	71
Esquema II.4: Esquema genérico de síntese de sais de imidazólio 28 e 29.....	72
Esquema II.5: Esquema de síntese de um sal de imidazólio 32, baseado em adição de Michael.	73
Esquema II.6: Síntese directa de sal bis(trifluorometilsulfonil)imidato 37 sem passar por halogeneto.	73
Esquema II.7: Novo esquema de síntese de sal de imidazólio funcionalizado com cadeia fluorada 41, usando agente alquilante 40.....	74
Esquema II.8: Reacção de fixação de CO ₂ por parte de epóxidos 43, catalisada pelo sal de guanidínio 44.....	76
Esquema II.9: Método de síntese de sal de guanidínio 49 a partir de ureia 46.....	78
Esquema II.10: Esquema de síntese de um sal de guanidínio 55 a partir de uma tioureia 52.....	78
Esquema II.11: Esquema de síntese de triflato de guanidínio 58, usando anidrido tríflico 56 e amina secundária.....	79
Esquema II.12: Esquema de síntese de um sal de guanidínio 64 a partir do tiocianato de potássio 59.....	79
Esquema II.13: Esquema de alquilação de uma guanidina protegida 69 com um álcool 70 usando o protocolo de Mitsunobu.....	80
Esquema II.14: Método simples de síntese de sais de guanidínio 76 e 78 a partir de cloreto de <i>N,N</i> -dialquilfosgeniminio 74.....	81
Esquema II.15: Método de síntese para obtenção de sais de imidazólio bromados.	101
Esquema II.16: Bromação olefinica de sal de imidazólio.....	102
Esquema II.17: Mecanismo reaccional relativamente à obtenção de brometo de 1-benzil-3-propilimidazólio.	150
Esquema II.18: Reacções de troca de anião para obtenção dos sais 2a e 3a obtidos a partir de brometo de 1-benzil-3-propilimidazólio 1a.....	151
Esquema II.19: Mecanismo reaccional para a síntese de cloretos de guanidínio.	172
Esquema II.20: Reacções de troca de anião cloreto por tetrafenilborato.....	177
Esquema II.21: Método sintético usado na síntese de sais de imidazólio bromados de elevada densidade. Bromação de posições olefinicas.	182
Esquema II.22: Método A - Reacção de bromação do cloreto de <i>N,N</i> -dimetil- <i>N',N',N'',N''</i> -tetraalilguanidínio 6a para a obtenção do cloreto de <i>N,N</i> -dimetil- <i>N',N',N'',N''</i> -tetraquis(2,3-dibromopropil) guanidínio 120.	183
Esquema II.23: Mecanismo reaccional da bromação de posição olefinica.....	183
Esquema II.24: Método B - Método de síntese modificado para a obtenção de sal de guanidínio de elevada densidade 11a, baseado em bromação de posições olefinicas posteriormente à reacção de troca de cloreto por anião não nucleófilo.....	186
Esquema II.25: Método C - Método de síntese modificado em relação a Esquema 24 – neste caso é usado o anião bis(trifluorometanosulfonil)imidato em vez de hexafluorofosfato.....	190
Esquema II.26: Método D - Método de síntese baseado na troca de cloreto por anião perfluorooctanoato, este tipo de estrutura organo-fluorada aparece associada a compostos de elevada densidade.....	195

Esquema II.27: Método E - Método de síntese para a obtenção de sal de elevada densidade 17a que não contenha número elevado de átomos de bromos na estrutura do catião.	196
Esquema II.28: Método F – Método de síntese de sal de guanidínio cujas cadeias substituintes possuem tamanhos assimétricos. Este tipo de característica estrutural aparece associada a compostos de baixo ponto de fusão.....	203

Capítulo III

Esquema III.1: Activação de aceitadores pro-Michael (Oxidação).	316
Esquema III.2: Reacção S_NAr	317
Esquema III.3: Substituições nucleófilas.....	317
Esquema III.4: Formação de base de Schiff 13.....	317
Esquema III.5: Electrófilos actuando como agentes acilantes.....	318
Esquema III.6: Activação metabólica da 3-dimetilaminopropilamina 18 e posterior reacção com proteína cutânea.....	318
Esquema III.7: Representação genérica de uma reacção em que intervém a butilamina.	344
Esquema III.8: Representação genérica de uma reacção em que intervém o cianoborohidreto de sódio.	346
Esquema III.9: Reacção contida no conjunto de treino mais próxima das reacções mal classificadas do conjunto de teste.....	388
Esquema III.10: Representação esquemática de reacção de formação de base de Schiff evidenciando a perda de duas ligações azoto-hidrogénio.	390
Esquema III.11: Representação de uma reacção de acilação resultando na formação de uma ligação carbono-azoto de amida, aparecendo aqui em evidência.	391
Esquema III.12: Reacção do tipo “f” classificada como sendo de tipo “b” onde também há a formação de uma ligação carbono-azoto de ureia.....	391
Esquema III.13: Reacção de substituição nucleofílica aromática “d” evidenciando a formação de ligação carbono-azoto sendo o carbono aromático.	391
Esquema III.14: Representação genérica de uma reacção em que intervém a butilamina.	399
Esquema III.15: Representação genérica de uma reacção em que intervém o cianoborohidreto de sódio.	421

Índice de Tabelas

Capítulo II

Tabela II.1-1: Descritores base usados no estudo de modelação de pontos de fusão de sais de imidazólio.	107
Tabela II.1-2: Continuação da apresentação dos descritores base usados no estudo de modelação de pontos de fusão de sais de imidazólio.	108
Tabela II.1-3: Continuação da apresentação dos descritores base usados no estudo de modelação de pontos de fusão de sais de imidazólio.	109
Tabela II.2-1: 246 descritores base usados no estudo da densidade de sais de imidazólio.	112
Tabela II.2-2: 246 descritores base usados no estudo da densidade de sais de imidazólio.	113
Tabela II.2-3: 246 descritores base usados no estudo da densidade de sais de imidazólio.	114
Tabela II.3-1: 92 descritores base usados no estudo do ponto de fusão de sais de guanidínio.	116
Tabela II.3-2: 92 descritores base usados no estudo do ponto de fusão de sais de guanidínio.	117
Tabela II.3-3: 92 descritores base usados no estudo do ponto de fusão de sais de guanidínio.	118
Tabela II.4: Composição dos conjuntos de treino e teste ao nível de distribuição de aniões e valores máximos e mínimos de ponto de fusão.	120
Tabela II.5: Composição do conjunto de treino e conjunto de teste ao nível de aniões e classes.	121
Tabela II.6: Distribuição de objectos em conjuntos de treino e teste segundo anião e gamas de densidade.	121
Tabela II.7: Distribuição de objectos por conjunto de treino e teste segundo anião e gamas de densidade.	122
Tabela II.8: Exemplo de uma tabela de contingência para 3 classes distintas.	130
Tabela II.9: Características dos modelos construídos para modelar o ponto de fusão de sais de imidazólio.	131
Tabela II.10: Comparação, entre modelos RF, SVM e AD aplicados ao conjunto de treino com e sem validação cruzada 5x.	133
Tabela II.11: Os 10 descritores moleculares mais importantes na construção de modelo <i>Random Forest</i> . Para cada descritor aparece representada medida de importância %IncMSE.	135
Tabela II.12: Comparação, ao nível de conjunto de treino e teste, entre modelo RF global e modelos RF individuais para cada tipo de anião.	137
Tabela II.13: Principais características dos modelos construídos.	139
Tabela II.14: Comparação, entre modelos RF, SVM e AD aplicados ao conjunto de treino com e sem validação cruzada 5x.	141
Tabela II.15: Tabela de contingência obtida por modelo SVM para conjunto de teste de 109 sais.	146
Tabela II.16: Resultados obtidos por modelo SVM para conjunto independente de teste de 109 compostos.	147
Tabela II.17: Previsões de classe estado físico/ponto de fusão obtidas para as duas estruturas alvo.	150
Tabela II.18: Dados reaccionais, espectroscópicos e análise elementar do brometo de 1-benzil-3-propilimidazólio (1a).	151
Tabela II.19: Dados reaccionais, espectroscópicos e análises elementares dos sais de tetrafluoroborato (2a) e trifluorometanosulfonato (3a).	152

Tabela II.20: Comparação entre a classe experimental e classes previstas pelos três modelos estudados (SVM, RF e AD) relativamente ao ponto de fusão (transição vítrea) e estado físico para as duas moléculas alvo neste estudo.	153
Tabela II.21: Informação adicional fornecida por modelo RF.....	154
Tabela II.22: Características dos três modelos construídos para modelar a densidade de sais de imidazólio.	155
Tabela II.23: Comparação do desempenho de três modelos distintos (RF, SVM e AD) relativamente a conjunto de treino com e sem validação cruzada 3x.	157
Tabela II.24: Os 10 descritores mais importantes na modelação da densidade de sais de imidazólio segundo modelo Random Forest.....	160
Tabela II.25: Resultados obtidos por modelo SVM na previsão de densidade para conjunto independente de teste de 34 sais de imidazólio.....	162
Tabela II.26: Modelos <i>Counter-Propagation</i> base para modelação do ponto de fusão de sais de guanidínio.....	164
Tabela II.27: Comparação entre modelos baseados numa rede CPG-NN e num ensemble de 5 redes CPG-NN.....	165
Tabela II.28: Resultados obtidos por modelo baseado em ensemble de 5 redes neuronais CPG em ensaio de validação cruzada 5x.	167
Tabela II.29: Dados reaccionais, espectroscópicos e análises elementares dos cloretos de guanidínio sintetizados.	174
Tabela II.30: Dados reaccionais, espectroscópicos e análises elementares dos cloretos de guanidínio sintetizados.	175
Tabela II.31-1: Dados reaccionais, espectroscópicos e análises elementares dos sais de guanidínio sintetizados cujo anião é o tetrafenilborato.	178
Tabela II.31-2: Continuação da apresentação dos dados reaccionais, espectroscópicos e análises elementares dos sais de guanidínio sintetizados cujo anião é o tetrafenilborato.....	179
Tabela II.32: Dados reaccionais, espectrais e análise elementar do composto intermediário 10a à síntese de sal de guanidínio de elevada densidade 18a.	187
Tabela II.33: Dados reaccionais e espectrais de sal de guanidínio bromado 11a, nos sinais relativos ao espectro de massa (+) representa os fragmentos correspondentes à combinação de isótopos maioritária.....	189
Tabela II.34: Dados reaccionais, espectrais e análise elementar do composto intermediário 12a à síntese de sal de guanidínio de elevada densidade 13a.	192
Tabela II.35: Dados reaccionais, espectrais e análise elementar de sal de guanidínio bromado 13a.....	193
Tabela II.36: Dados reaccionais, espectrais e análise elementar de sal de guanidínio 14a contendo anião perfluorooctanoato.....	197
Tabela II.37: Dados reaccionais, espectrais e análise elementar de sal de guanidínio 15a.	198
Tabela II.38: Dados reaccionais, espectrais e análise elementar de sal de bis(trifluorometanosulfonil)imidato 16a.....	200
Tabela II.39: Dados reaccionais, espectrais e análise elementar de sal tetrabromado 17a.	201
Tabela II.40: Dados espectroscópicos referentes à amina secundária 18c.	203
Tabela II.41: Dados espectroscópicos referentes a sal de guanidínio 18a contendo grupos substituintes diferenciados em termos de tamanho.	204
Tabela II.42: Comparação ao nível do ponto de fusão e densidade entre compostos intervenientes no método B de síntese.	206
Tabela II.43: Comparação ao nível do ponto de fusão e densidade entre compostos intervenientes no método C de síntese.	207

Tabela II.44: Comparação ao nível do ponto de fusão e densidade entre compostos intervenientes no método D de síntese.	208
Tabela II.45: Comparação ao nível do ponto de fusão e densidade entre compostos intervenientes no método E de síntese.	210
Tabela II.46: Transição vítrea do cloreto de guanidínio 18a obtido pelo método F de síntese.....	211
Tabela II.47-1:Dados reaccionais, espectrais e análise elementar do sal baseado no anião tiocianato 19a e dicianamida 20a.	215
Tabela II.47-2:Dados reaccionais, espectrais e análise elementar do sal baseado no anião tiocianato 19a e dicianamida 20a.	216
Tabela II.48: Dados reaccionais, espectroscópicos e análise elementar do sal baseado no anião dicianamida 21a e tiocianato 22a.	218
Tabela II.49: Dados reaccionais, espectrais e análises elementares relativamente ao sal contendo anião <i>para</i> -toluenosulfonato 23a.	220
Tabela II.50: Comparação entre diversos sais baseados no catião de <i>N,N,N',N''</i> -tetrametil- <i>N',N''</i> -dioctadecilguanidínio relativamente ao ponto de fusão.	221
Tabela II.51: Comparação entre diversos sais baseados no catião <i>N,N</i> -dimetil- <i>N',N',N'',N''</i> -tetraalilguanidínio relativamente ao ponto de fusão.	222
Tabela II.52: Comparação entre dois sais baseados no catião <i>N,N</i> -dimetil- <i>N',N''</i> -dietil, <i>N',N''</i> -bis(2-metilalil)guanidínio ao nível da transição para o estado líquido.	223
Tabela II.53: Cloreto de <i>N,N</i> -dimetil- <i>N',N''</i> -dialil- <i>N',N''</i> -dioctilguanidínio e respectiva transição vítrea.	224
Tabela II.54: Comparação entre diversos sais baseados no catião <i>N,N</i> -dimetil- <i>N',N',N'',N''</i> -tetrahexilguanidínio ao nível da transição para o estado líquido.	225
Tabela II.55: Comparação entre diversos sais baseados no catião <i>N,N</i> -dimetil- <i>N',N''</i> -dihexil- <i>N',N''</i> -bis(2-etilhexil)guanidínio ao nível da transição para o estado líquido.	226

Capítulo III

Tabela III.1-1: Descritores de ligações usados no estudo a) de mecanismo reaccional associado a determinado composto.....	334
Tabela III.1-2: Continuação da apresentação dos descritores de ligações usados no estudo a) de mecanismo reaccional associado a determinado composto.	335
Tabela III.1-3: Continuação da apresentação dos descritores de ligações usados no estudo a) de mecanismo reaccional associado a determinado composto.	336
Tabela III.1-4: Continuação da apresentação dos descritores de ligações usados no estudo a) de mecanismo reaccional associado a determinado composto.	337
Tabela III.2: 11 descritores físico-químicos calculados pelo programa Petra e suas especificações.....	338
Tabela III.3: Tipos e dimensões de MOLMAPs usados no estudo de mecanismo reaccional de composto. Em todos os casos foram usados descritores do tipo 10_3.....	340
Tabela III.4: Tipos e dimensões de MOLMAPs usados no estudo de mecanismo reaccional associado a uma determinada reacção.	341
Tabela III.5: Dimensões de MOLMAP usados no estudo de reactividade na presença de butilamina.....	341

Tabela III.6: Dimensões de MOLMAPs usados no estudo de reactividade na presença de cianoborohidreto de sódio.....	342
Tabela III.7: Composição dos conjuntos de treino, validação e teste relativamente aos diferentes tipos de mecanismos reaccionais.....	343
Tabela III.8: Características dos modelos <i>Random Forest</i> construídos, tendo em conta o tipo de MOLMAP, suas dimensões e método de obtenção de pseudocompostos que não reagem.	352
Tabela III.9: Características dos MOLMAPs usadas na construção de modelos <i>Random Forest</i> , variando o tipo e as dimensões dos MOLMAP.	355
Tabela III.10: Capacidade do mapa auto-organizativo produzido com conjunto de 59 descritores topológicos/físico-químicos para discriminar ligações químicas iguais ou semelhantes mas com ambiente químico diferente.	362
Tabela III.11: Capacidade do mapa auto-organizativo produzido com conjunto de 27 descritores físico-químicos para discriminar ligações químicas.	364
Tabela III.12: Capacidade do mapa auto-organizativo produzido com conjunto de 45 descritores topológicos para discriminar ligações químicas.....	366
Tabela III.13: Comparação entre os diversos modelos <i>Random Forest</i> treinados para prever o mecanismo a partir do substrato representado com MOLMAPs de diversas dimensões baseados em 45 propriedades topológicas de ligação química.	367
Tabela III.14: Comparação, relativamente à concordância, entre os diversos modelos <i>Random Forest</i> treinados para prever o mecanismo a partir do substrato representado com MOLMAPs de diversas dimensões baseados em 27 propriedades físico-químicas de ligação química.	368
Tabela III.15: Comparação, relativamente à Concordância, entre os diversos modelos <i>Random Forest</i> treinados para prever o mecanismo a partir do substrato representado com MOLMAPs de diversas dimensões baseados em 59 propriedades físico-químicas/topológicas de ligação química.	368
Tabela III.16: Resultados obtidos para a classificação de compostos de acordo com o mecanismo reaccional (a-f) nos conjuntos de treino, de validação e de teste.	369
Tabela III.17: Tabela de contingência relativamente aos resultados obtidos com o conjunto de teste de 45 moléculas.	370
Tabela III.18-1: Compostos pertencentes ao conjunto de teste de 45 moléculas e respectivas classes de mecanismo reaccional, de acordo com a literatura e classe prevista por modelo RF.	372
Tabela III.18-2: Continuação da apresentação dos compostos pertencentes ao conjunto de teste de 45 moléculas e respectivas classe de mecanismo reaccional, de acordo com a literatura, e classe prevista por modelo RF.	373
Tabela III.18-3: Continuação da apresentação dos compostos pertencentes ao conjunto de teste de 45 moléculas e respectivas classe de mecanismo reaccional, de acordo com a literatura, e classe prevista por modelo RF.	374
Tabela III.19: Os 10 descritores mais importantes seleccionados por modelo <i>Random Forest</i> e respectiva medida de importância.	377
Tabela III.20: Comparação de quatro modelos baseados em redes neuronais de <i>Kohonen</i> relativamente à capacidade de classificação, de acordo com o mecanismo reaccional (a-f), referente ao conjunto de treino e conjunto independente de validação.	380
Tabela III.21: Comparação, relativamente ao valor de concordância, entre os diversos modelos <i>Random Forest</i> construídos com MOLMAPs de diversas dimensões baseados em 59 propriedades físico-químicas de ligação química.	384
Tabela III.22: Optimização do valor de <i>mtry</i> usando MOLMAPs 15x15. Comparação entre diversos valores de <i>mtry</i> é feita através do valor de Concordância.	385

Tabela III.23: Modelo <i>Random Forest</i> <i>mtry</i> = 25 usando MOLMAPs 15x15, estudo detalhado relativamente à capacidade de classificação, de acordo com o mecanismo reaccional (a-f), referente ao conjunto de treino, conjunto independente de validação e conjunto de teste.....	386
Tabela III.24: Tabela de contingência relativamente aos resultados obtidos com o conjunto de teste de 45 reacções.	387
Tabela III.25: Os dez descritores segundo o modelo <i>Random Forest</i> – MOLMAP 15x15, <i>mtry</i> = 25 e respectivas medidas de importância.....	393
Tabela III.26: Capacidade do mapa auto-organizativo para discriminar ligações representadas por um conjunto de 11 descritores físico-químicos.....	398
Tabela III.27: Comparação, relativamente ao valor de Concordância, entre os diversos modelos <i>Random Forest</i> construídos com MOLMAPs de diversos tipos e dimensões.	400
Tabela III.28: Comparação, relativamente ao valor de concordância, entre os diversos modelos <i>Random Forest</i> construídos com MOLMAPs de diversas dimensões.	404
Tabela III.29: Optimização do valor de <i>mtry</i> usando MOLMAPs 29x29 e 20x20. Comparação entre diversos valores de <i>mtry</i> é feita através do valor de concordância.	404
Tabela III.30: Valores de Concordância, especificidade e sensibilidade obtidos por melhor modelo RF para conjuntos de treino, validação e teste.....	405
Tabela III.31-1: Lista de compostos testados <i>in vitro</i> na presença de péptido contendo lisina.....	407
Tabela III.31-2: Continuação da lista de compostos testados <i>in vitro</i> na presença de péptido contendo lisina.	408
Tabela III.31-3: Continuação da lista de compostos testados <i>in vitro</i> na presença de péptido contendo lisina.	409
Tabela III.32: Tipos de ligação representadas em árvore de decisão da Figura III.34.	415
Tabela III.33: Os dez descritores mais importantes segundo melhor modelo RF – 1000 árvores, <i>mtry</i> = 50 obtido com MOLMAPs 20x20 do tipo 1_0 e respectivos valores de importância.	416
Tabela III.34: Comparação, relativamente a valores de Concordância, entre modelos baseados em redes de <i>Kohonen</i> 20x20 – baseados em MOLMAPs 29x29 do tipo 1_0 e 10_1 – selecção de 64 descritores vs MOLMAP completo.	417
Tabela III.35: Comparação, relativamente a valores de concordância, entre modelos baseados em redes de <i>Kohonen</i> de diferentes dimensões - modelos baseados em selecção de 64 descritores provenientes de MOLMAP 29x29 do tipo 10_1.	418
Tabela III.36: Resultados obtidos, relativamente à concordância, especificidade e sensibilidade, para conjuntos de treino, validação e teste com melhor modelo baseado em rede neuronal de <i>Kohonen</i>	418
Tabela III.37: Valores de Concordância obtidos por modelos SVM variando o parâmetro C de complexidade – usando selecção de 64 descritores provenientes de MOLMAP 29x29 do tipo 10_1.....	420
Tabela III.38: Resultados obtidos, relativamente à concordância, especificidade e sensibilidade, para conjuntos de treino, validação e teste com modelo baseado em algoritmo <i>Support Vector Machines</i>	420
Tabela III.39: Concordâncias relativamente a conjunto de treino, ensaio <i>Out of Bag</i> e conjunto de validação referentes a Modelos <i>Random Forest</i> baseados em diferentes tipos e dimensões de MOLMAP.	423
Tabela III.40: Valores de Concordância obtidos com diferentes valores de <i>mtry</i> de forma a otimizar modelo de previsão de reactividade na presença de cianoborohidreto de sódio. Conjunto de treino <i>Out of Bag</i> e conjunto de validação.....	423

Tabela III.41: Resultados obtidos, relativamente à concordância, especificidade e sensibilidade por modelo RF seleccionado para conjuntos de treino, validação e teste.	424
Tabela III.42-1: Lista de compostos gerados a partir de informação recolhida na literatura relativamente à reactividade de determinadas funcionalidades na presença de cianoborohidreto de sódio.	426
Tabela III.42-2: Continuação da lista de compostos gerados a partir de informação recolhida na literatura relativamente à reactividade na presença de cianoborohidreto de sódio.....	427
Tabela III.43: Tipos de ligação representadas em árvore de decisão da Figura III.38..	429
Tabela III.44: Os dez descritores mais importantes segundo melhor modelo RF – 1000 árvores, <i>mtry</i> = 75 obtido com MOLMAPs 15x15 do tipo 1_0 e respectivos valores de importância.	430

Nota Introdutória:

O trabalho descrito ao longo desta dissertação está dividido em quatro capítulos. O capítulo II descreve o estudo de propriedades físico-químicas de líquidos iónicos numa perspectiva em simultâneo Químio-Informática e Química Orgânica. Nesse estudo foram construídos modelos de previsão de pontos de fusão e densidade que posteriormente foram validados através de comparação entre previsões obtidas pelos modelos e valores dessas propriedades exibidas pelos respectivos compostos sintetizados e/ou compostos recolhidos da literatura e reunidos em conjunto independente de teste. Foram igualmente sintetizados novos líquidos iónicos com o objectivo de encontrar sais líquidos à temperatura ambiente e/ou com elevada densidade mas cujas características estruturais associadas a catião e/ou anião não permitiram a obtenção de previsões por parte de modelos computacionais, o motivo para essa impossibilidade deveu-se à escassez de exemplos na literatura de compostos com características estruturais concretas semelhantes aos que se pretendiam sintetizar. Na impossibilidade prever propriedades para esses compostos, a sua síntese baseou-se em informação experimental recolhida na literatura sobre características estruturais que levariam à obtenção de líquidos iónicos com as propriedades pretendidas. Do trabalho desenvolvido neste capítulo resultaram dois artigos, um já publicado “G. V. S. M. Carrera, L.C. Branco, J. Aires-de-Sousa, C.A. M. Afonso, Exploration of quantitative structure-property relationships (QSPR) for the design of new guanidinium ionic liquids, *Tetrahedron* **2008**, 64(9), 2216-2224”. Outro artigo encontra-se em fase de submissão: “G. V. S. M. Carrera, L. C. Branco, R. Frade, J. Aires-de-Sousa, C. A. M. Afonso, Physico-chemical properties of new functionalized guanidinium based ionic liquids and molten salts, **2009**”. O capítulo III reporta-se em exclusivo ao estudo Químio-Informático de reactividade química de compostos químicos e classificação segundo mecanismo reaccional de compostos e reacções. Este estudo foi feito a partir da aprendizagem automática de padrões mecanísticos e de reactividade presentes em bases de dados extensas de reacções químicas. Esses padrões foram codificados através de metodologia MOLMAP desenvolvida no grupo onde foi realizado este trabalho. Resultou do trabalho desenvolvido neste capítulo um artigo já aceite para publicação: G. V. S. M. Carrera, S. Gupta, J. Aires-de-Sousa, Machine learning of chemical reactivity from databases of organic reactions, *Journal of Computer-Aided Molecular Design* **2009**.

Por sua vez, o capítulo I é uma compilação de toda a metodologia Químio-Informática transversal aos capítulos II e III, referindo-se essencialmente à descrição métodos computacionais de aprendizagem automática e à caracterização de descritores moleculares/reacionais. Finalmente, o capítulo IV é um Glossário estando aí explicados

alguns termos que aparecem ao longo da tese mas que não foram aprofundados por não se enquadrarem com os objectivos propostos.

Capítulo I

Métodos de aprendizagem automática e descritores moleculares

I - 1: Métodos de Aprendizagem Automática

Na obtenção de modelos para prever propriedades físico-químicas de líquidos iónicos, reactividade química de compostos, ou o mecanismo reaccional associado a um composto ou a uma reacção, foram usados métodos de aprendizagem automática. Estes consistem em algoritmos e técnicas que permitem os computadores aprender de forma automática, extraindo informação de um determinado conjunto de dados. Em concreto, vão ser focados métodos baseados em redes neuronais de *Counter-Propagation* e *Kohonen*, Árvore de Decisão, *Random Forest* e *Support Vector Machines*.

I-1.1: Redes Neuronais

Na tentativa de simular a arquitectura e a forma de processamento do cérebro humano foram desenvolvidos variados modelos e algoritmos matemáticos denominados redes neuronais artificiais. Para melhor compreensão deste conceito matemático importa conhecer *à priori* as características fundamentais do fenómeno biológico de processamento de informação por parte do cérebro humano. O cérebro humano é essencialmente uma rede de *ca* de 10^{10} células, denominadas neurónios, que interagem entre si através das respectivas interfaces, denominadas sinapses (**Figura I.1**).

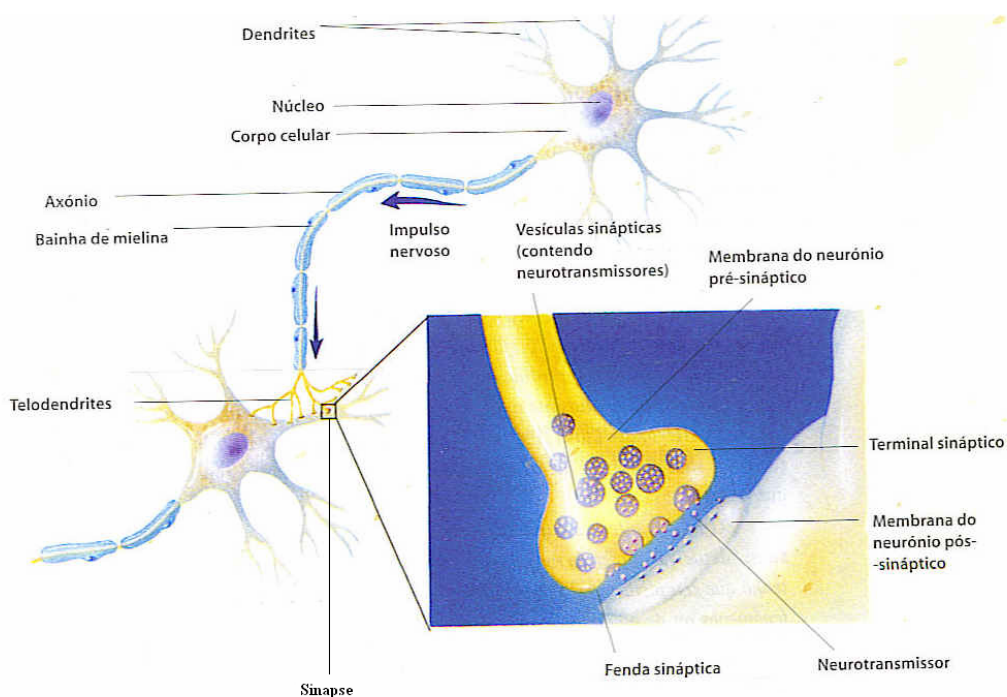


Figura I.1: Estrutura de um neurónio e interação com outro neurónio através de sinapses. ¹

Essencialmente, as dendrites recebem sinais químicos, que são acumulados no corpo celular até atingir certo limite, altura em que o neurónio produz um impulso eléctrico que é transmitido através do axónio atingindo as telodendrites e estimulando a libertação de neurotransmissores, responsáveis pela transmissão de sinal às dendrites do neurónio vizinho. O local onde ocorre a transmissão do sinal, a sinapse, tem também como função a constituição de uma barreira que modela a intensidade do sinal x_i durante a transmissão, estando a força sináptica (ou peso sináptico) w_j na génese dessa modelação. A força sináptica vai ser alterada ao longo do tempo adaptando-se a um determinado problema, constituindo esta adaptação a essência da aprendizagem. Relativamente ao processamento de determinado sinal através da rede neuronal, este é feito em paralelo já que determinada informação é canalizada simultaneamente por diversos neurónios. Esta forma em paralelo de processar a informação é substancialmente mais eficaz do que o processamento sequencial.

O conceito de rede neuronal artificial apareceu pouco tempo depois do aparecimento dos primeiros computadores. A primeira simulação computacional de redes neuronais foi feita por Rochester e seus colaboradores em 1956 marcando assim o aparecimento do conceito de inteligência artificial. O passo seguinte na evolução das redes neuronais artificiais foi dado por Rosenblatt que em 1959 desenvolveu o *perceptrão*, que consistia numa estrutura neuronal de uma camada que produzia *output* escalar 1 ou -1. Esta estrutura apenas permitia resolver tipos de problemas bastante restritos, caso de problemas que fossem linearmente separáveis e que correspondessem à condição XOR - condição verdadeira se uma e só uma das variáveis é verdadeira. Claramente o conceito de rede neuronal artificial estava muito longe de simular o funcionamento do cérebro humano, e durante mais de duas décadas houve actividade reduzida no desenvolvimento de redes neuronais.

Em 1982 apareceu um contributo relevante por Hopfield, que introduziu funções de transferência não lineares para o processamento de um determinado sinal x_i sujeito a pesos sinápticos w_j . Desta forma cada neurónio adquiriu um carácter não linear e as redes neuronais adquiriram uma flexibilidade que não existia nos modelos prévios. O algoritmo desenvolvido passou a chamar-se rede de Hopfield. No mesmo ano Kohonen² desenvolveu um novo conceito de redes baseado na topologia da informação em que um objecto seria mapeado de acordo com um vector de *input* ou sinal que o representava. No final objectos com sinal semelhante seriam mapeados em neurónios próximos. As redes de *Kohonen* incorporavam ainda dois novos conceitos. O primeiro foi o conceito de selecção de neurónio vencedor. O segundo, talvez mais importante, consiste no ajuste dos pesos do neurónio vencedor de forma a tornar o neurónio ainda mais capaz de ser activado pelo sinal do objecto apresentado. Para isso foi definido um algoritmo de ajuste de pesos. Usando os mesmos conceitos introduzidos por Kohonen, Hecht-Nielsen em 1987 desenvolve um novo tipo de rede, denominada rede neuronal de *Counter-propagation*,² que é idêntica à rede de *Kohonen* quer no conceito da topologia da informação, quer no conceito de neurónio vencedor, quer no algoritmo utilizado para

ajustar os pesos do neurónio vencedor. Aqui a grande inovação consiste na introdução de uma nova camada, a camada de *output*, em adição à camada de neurónios de *Kohonen*. A camada de *output*, à semelhança da camada de *Kohonen*, vai ter os respectivos pesos alterados durante a fase de treino ou aprendizagem, o algoritmo usado para esse efeito é idêntico ao que foi introduzido por Teuvo Kohonen.

Já anteriormente havia sido desenvolvido um outro tipo de algoritmo para ajuste dos pesos sinápticos durante a fase de treino aplicado a redes contendo diversas camadas de neurónios, primeiro por Parker em 1985 e posteriormente, num estudo independente, por Rumelhart, Hinton e Williams em 1986. Posteriormente veio a reconhecer-se que este algoritmo havia sido implementado muito mais cedo por Werbos na sua tese de doutoramento na Universidade de Harvard em 1974. Esta arquitectura multi-camada associada ao algoritmo de ajuste de pesos ficou conhecida como rede de Retropropagação (*Backpropagation*) visto os pesos serem ajustados a partir da camada de *output* até à primeira camada. Mais concretamente, os desvios entre valores previstos (*output*) e valores experimentais na camada de *output* vão servir para corrigir os pesos que ligam os neurónios da camada de *output* aos neurónios da camada imediatamente acima. O procedimento para a correcção dos pesos vai sendo repetido para as camadas mais próximas da camada de *input*. Desta forma propaga-se a correcção dos erros em sentido contrário (retropropagação) ao da obtenção de previsões pelo modelo. Genericamente, numa rede neuronal *backpropagation*, o *input* é introduzido nos neurónios da 1ª camada. Cada valor é transmitido ao neurónio seguinte, multiplicando pelo peso específico associado à ligação entre neurónios. Os neurónios que recebem sinais de vários neurónios, somam-nos e transformam a soma através de uma função de transferência. O resultado (*output* do neurónio) poderá ser a resposta final a um problema ou o *input* para neurónios na camada imediatamente abaixo (Figura I.2).

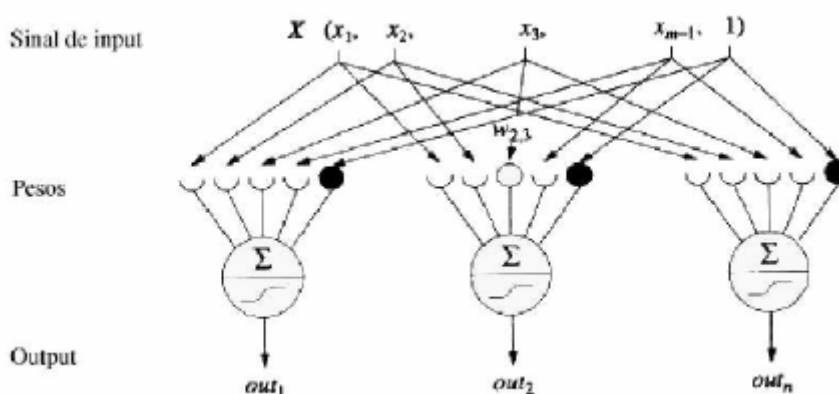


Figura I.2: Representação da incorporação e processamento de um determinado sinal de *input* por parte de uma rede neuronal *Backpropagation*.

Os valores de *input* podem ser números reais, preferencialmente com valores compreendidos entre 0 e 1, ou -1 e 1. O tipo de problemas resolvidos por redes neuronais é bastante variado, podendo dividir-se em cinco grandes grupos:

Auto-associação: Neste caso o sistema é capaz de reconstruir um padrão correcto mesmo que o padrão aprendido seja incompleto ou corrupto.

Hetero-associação: Aqui, o sistema vai produzir uma correspondência um a um entre membros de dois conjuntos distintos.

Classificação: Nesta abordagem o sistema vai atribuir a cada um dos objectos em estudo uma determinada classe, sendo agrupados em subconjuntos ou *clusters* de objectos. Essa classificação é baseada numa série de características que constituem o sinal de *input* de cada objecto.

Transformação ou mapeamento de um espaço vectorial com determinado número de dimensões, num novo espaço vectorial com o mesmo número de dimensões ou número mais reduzido, sendo também possível a conversão para um número de dimensões mais elevada, muito embora este último caso seja muito mais raro.

Modelação: Que consiste na busca de um procedimento ou função analítica que obtenha um determinado vector de *output* a partir de um vector de *input*. As redes neuronais não vão produzir uma expressão matemática que faça essa conversão, vão sim adaptar os pesos sinápticos associados ao processamento do sinal.

Neste trabalho foram utilizadas redes neuronais de *Counter-propagation* e *Kohonen* sendo descrito seguidamente o funcionamento destes dois tipos de redes.

I-1.1.1: Redes Neuronais de *Counter-Propagation*

Introduzidas por Hecht-Nielsen em 1987² as redes neuronais de *Counter-Propagation* (CPG-NN) têm sido usadas numa grande variedade de aplicações, nomeadamente no estudo de actividade inibitória da enzima HIV-1 transcriptase reversa,³ sendo analisada para o efeito um conjunto de compostos estruturalmente similares cuja actividade inibitória era conhecida. Redes CPG-NN têm também sido usadas para estudar afinidade de ligação entre a trombina, uma protease responsável pela regulação do aparecimento da trombose, e determinados compostos,⁴ estas redes têm também sido usadas na classificação de reagentes⁵ segundo o seu papel em reacções, na previsão de actividade anti-oxidante de compostos fenólicos,⁶ na previsão de pontos de fusão de líquidos iónicos^{7,8}, na modelação

da mudança de coloração associada a tintas em função da concentração de determinados óxidos e modelação do estado reaccional a dado instante associado a determinado sistema e troca entre variáveis de *input* e *output*,⁹ redes neuronais CPG foram também usadas na previsão de índices de Kovats,¹⁰ previsão de modos de acção tóxica em compostos fenólicos,¹¹ classificação de ácidos nucleicos e sequências proteicas,¹² modelação de mutagenicidade de aminas aromáticas e hetero-aromáticas,¹³ modelação do coeficiente de partição n-octanol/água, momento dipolar e refractividade molecular de compostos aromáticos,¹⁴ previsão de propriedades físico-químicas de solventes orgânicos,¹⁵ modelação de diversas actividades de benzodiazepinas e carboquinonas.¹⁶

Uma rede neuronal de *Counter-Propagation* é composta, genericamente, por duas camadas, uma camada de *input* e outra de *output*. Cada camada consiste numa grelha bidimensional de neurónios. Cada neurónio na camada de *input* tem tantos pesos (valores numéricos) como o número de variáveis de *input* (descritores) e cada neurónio de *output* tem tantos pesos como o número de propriedades a prever. A rede aprende ajustando os pesos durante a fase de treino, no qual os objectos do conjunto de treino são repetidamente submetidos à rede (**Figura I.3**).

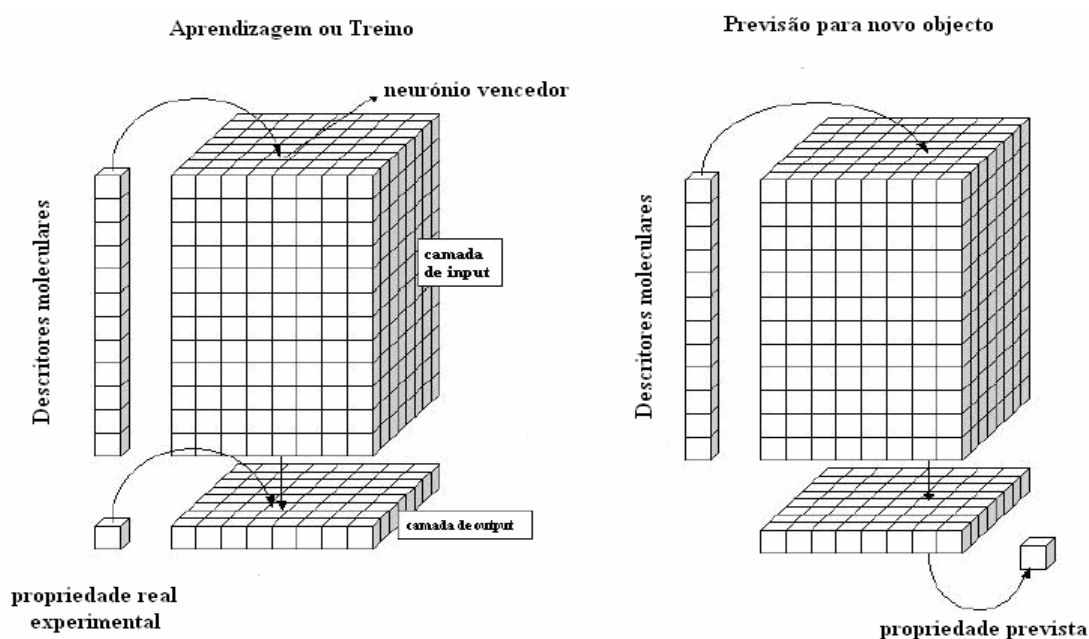


Figura I.3: Treino e previsão numa rede neuronal *Counter-Propagation* que aprende relações entre descritores moleculares (camada de *input*) e propriedade a modelar (camada de *output*) durante a fase de treino. Na fase de previsão, vão ser obtidas previsões para novos objectos.

Antes do treino, os pesos são gerados aleatoriamente. Durante o treino, e de acordo com o algoritmo usado neste trabalho, cada objecto activa o neurónio com os pesos de *input* mais

semelhantes aos descritores desse mesmo objecto – esse é considerado o neurónio vencedor. Essa semelhança é avaliada em termos da distância Euclidiana:

$$Q_i = (x_1 - w_{j1})^2 + (x_2 - w_{j2})^2 + \dots + (x_n - w_{jn})^2$$

$$j = 1, 2, \dots, s$$

Equação I.1

em que s é o número de neurónios da grelha 2D, x_i é o valor do descritor i do vector de *input* n dimensional, w_{ji} o peso correspondente na rede neuronal. Assim o neurónio vencedor é aquele que tiver o valor Q_j mais baixo. Posteriormente, os pesos do neurónio vencedor são ajustados de maneira a tornarem-se ainda mais semelhantes ao vector de *input* (descritores). Também os neurónios vizinhos são corrigidos de maneira a aproximarem-se do objecto de *input* (descritores moleculares),

$$\begin{aligned} w_{j1}^{(novo)} &= w_{j1}^{(antigo)} + \eta(t).a(d_c - d_j).(x_1 - w_{j1}^{(antigo)}) \\ w_{j2}^{(novo)} &= w_{j2}^{(antigo)} + \eta(t).a(d_c - d_j).(x_2 - w_{j2}^{(antigo)}) \\ &\dots \\ w_{jn}^{(novo)} &= w_{jn}^{(antigo)} + \eta(t).a(d_c - d_j).(x_n - w_{jn}^{(antigo)}) \end{aligned}$$

$$j = 1, 2, \dots, n$$

Equação I.2

com o grau de correcção dependente da distância ao neurónio vencedor ($d_c - d_j$) e de uma função do número de vezes que o conjunto de treino foi submetido à rede neuronal ($\eta(t)$). Nesta expressão c representa o neurónio vencedor, j o neurónio vizinho. A função $\eta(t)$ vai ter valor decrescente à medida que o treino vai decorrendo:

$$\eta(t) = (a_{\max} - a_{\min}) \frac{t_{\max} - t}{t_{\max} - 1} + a_{\min}$$

Equação I.3

sendo a_{max} , a_{min} e t_{max} constantes correspondendo respectivamente ao valor máximo e mínimo que a função vai ter e ao número máximo ou total de objectos que vão ser apresentados durante o treino. A função $a(d_c - d_j)$ vai ter os seguintes perfis, atingindo o valor máximo para o neurónio vencedor.

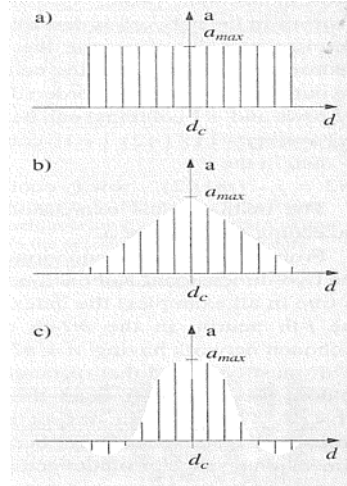


Figura I.4: Perfis mais comuns da função $a(d_c - d_j)$, com valores máximos para o neurónio vencedor. Três tipos de funções diferentes: a) constante, b) triangular, c) chapéu mexicano.²

Da mesma maneira, os correspondentes pesos de *output* são corrigidos para se aproximarem do valor experimental das propriedades que se pretendem prever; os pesos dos neurónios vizinhos na camada de *output* são também corrigidos como na camada de *input*. Este método, que vai corrigindo o valor de *output* durante o treino é um método supervisionado. A correcção dos pesos da camada de *output* é semelhante à da camada de *input*, representada pela seguinte equação:

$$\begin{aligned}
 C_{j1}^{(novo)} &= C_{j1}^{(antigo)} + \eta(t).a(d_c - d_j).(y_1 - C_{j1}^{(antigo)}) \\
 C_{j2}^{(novo)} &= C_{j2}^{(antigo)} + \eta(t).a(d_c - d_j).(y_2 - C_{j2}^{(antigo)}) \\
 &\dots \\
 C_{jm}^{(novo)} &= C_{jm}^{(antigo)} + \eta(t).a(d_c - d_j).(y_m - C_{jm}^{(antigo)})
 \end{aligned}$$

$j=1,2,\dots, m$

Equação I.4

sendo C_{ji} o peso da camada de *output* e y_i o valor experimental da propriedade i do vector de *output* m dimensional. Posteriormente a rede de *Counter-Propagation* treinada está em condições de produzir previsões para a propriedade modelada, para novos objectos. Para isso, à semelhança da fase de treino, vai ser encontrado o neurónio cujos pesos da camada de *input* são mais semelhantes às propriedades do novo objecto, a essa localização vai corresponder um determinado valor na camada de *output*, a previsão para esse objecto. Com este tipo de redes, para além da modelação de uma determinada propriedade em função de uma série de variáveis de *input* associadas a um objecto, obtém-se também a localização de um objecto num mapa em função do seu perfil de pesos de *input*. Pode também ser definido o grau de correlação entre níveis de *input*, ou entre determinado nível de *input* e determinado nível de *output*.

I-1.1.2: Redes Neurais de *Kohonen*

Introduzidas por Teuvo Kohonen em 1982,² as redes neuronais de *Kohonen* já foram utilizadas para classificar reacções químicas segundo o seu mecanismo reaccional,^{17,18} para classificar reacções metabólicas,¹⁹ para seleccionar conjunto de treino tendo em conta a variabilidade no espaço de *input*,²⁰ para a classificação geográfica de petróleos,²¹ para obtenção de descritores MOLMAP representando o perfil das ligações químicas de um composto,¹⁸ para a classificação da interacção de compostos com a glicoproteína P (P-gp) transmembranar, responsável pela biodisponibilidade de fármacos no interior da célula,²² separar moléculas de actividade biológica distinta, desenvolvimento de um teste na pesquisa de moléculas activas relativamente a uma determinada actividade biológica, avaliação da qualidade dos descritores a usar na construção de modelo baseado em método supervisionado. Estas aplicações e respectivas referências podem ser encontradas no artigo de revisão de Gasteiger *et al.*²³

Outras aplicações igualmente interessantes, como é o caso do agrupamento de sequências proteicas em diferentes famílias, comparação de bases de dados de estruturas químicas tanto na semelhança como na diversidade molecular e padrão na comparação e avaliação de potencial electrostático contido em superfície de molécula alvo, podem ser encontradas em artigo de revisão de Niculescu²⁴ e referências aí citadas.

As redes neuronais de *Kohonen* estão constituídas por uma mono-camada de neurónios, semelhante à camada de *input* das redes CPG. O procedimento de treino é semelhante ao da rede CPG, em que a um objecto submetido à rede de *Kohonen* vai corresponder um neurónio, o neurónio vencedor, ou seja, aquele cujos n pesos vão ser mais semelhantes às n variáveis do objecto submetido em termos de distância Euclidiana (**Equação I.1**). Durante o treino os pesos do neurónio vencedor vão ser corrigidos de forma a ficar mais próximos das variáveis de *input* do objecto; também os pesos dos neurónios vizinhos vão ser modificados de acordo com a distância ao neurónio vencedor e ao número

de objectos que já foram submetidos à rede até essa altura (**Equação I.2**). A principal diferença está relacionada com o facto de que as rede de *Counter-Propagation* possuem duas camadas, uma de *input* e outra de *output*, sendo os pesos dos neurónios de ambas as camadas modificados durante o treino, enquanto que as redes de *Kohonen* possuem apenas uma camada de *input*, cujos pesos vão ser alterados durante o treino (**Figura I.5**).

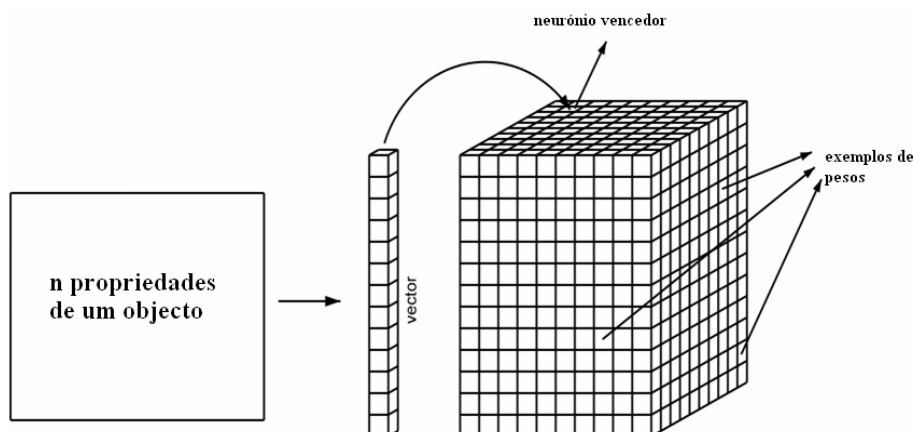


Figura I.5: Processo de treino de uma rede de *Kohonen* em que um objecto n dimensional está a ser submetido à rede, levando à correcção de uma série de pesos do neurónio vencedor e neurónios vizinhos.

Após a fase de treino, novos objectos podem ser submetidos à rede sendo obtida uma localização (localização do neurónio vencedor) para esse objecto que pode ser contrastada com a localização de outros objectos, sendo possível avaliar características em comum com objectos mapeados na mesma zona. Estas redes permitem ainda a classificação de um objecto de acordo com o problema em estudo.

Note-se que as redes de *Kohonen* aprendem de forma não supervisionada, isto é, o mapeamento dos objectos na rede é feito exclusivamente com base nas variáveis de *input* do objecto. Quando estas redes são usadas para classificação, o algoritmo inspecciona que objectos foram mapeados em cada neurónio e só nessa altura atribui uma classe ao neurónio, a classe dos objectos que o activaram.

Tanto nas redes neuronais de *Kohonen* como nas de *Counter-Propagation*, as camadas de neurónios (**Figura I.3** e **Figura I.5**), ainda que sejam representadas de forma quadrangular (ou rectangular), apresentam uma topologia toroidal – o lado esquerdo e direito da rede estão unidos, tal como a parte inferior e superior, tal como aparece representado na **Figura I.6**.

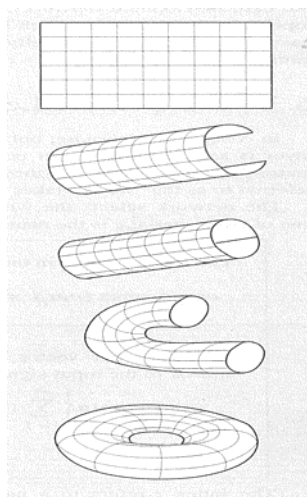


Figura I.6: A aparente forma quadrangular (rectangular) das redes neuronais de *Kohonen* e de *Counter-Propagation* representa na realidade um topologia toroidal.²

I-1.2: Árvores de Decisão

É um método de classificação e regressão conceptualmente simples, baseado na partição sequencial de um conjunto de objectos em subconjuntos progressivamente mais pequenos e mais homogéneos relativamente à propriedade quantitativa ou qualitativa em estudo, em que cada partição é obtida através de uma condição lógica baseada numa característica (descriptor). O uso de árvores para resolver problemas de regressão foi iniciado por Morgan e Sonquist no contexto do programa AID (Automatic Interaction Detection) desenvolvido na Universidade do Michigan no início da década de 60.²⁵ Já o algoritmo para classificação foi desenvolvido no início da década de 70 por Morgan e Messenger. O conceito de árvores de decisão evoluiu consideravelmente a partir de 1973 quando Breiman e Friedman, independentemente um do outro, desenvolveram novos métodos baseados em árvores de classificação.²⁶ Mais tarde, Stone e Olshen contribuíram significativamente para a evolução deste algoritmo em termos metodológicos e teóricos respectivamente. Este conjunto de ideias confluiu na criação do algoritmo CART (Classification and Regression Trees) em 1984.²⁷ Árvores de decisão, aliando capacidade de previsão, simplicidade e interpretabilidade dos modelos obtidos, constituem ferramentas de trabalho muito úteis tendo sido utilizadas em inúmeras aplicações tal como o estudo de inibidores da transcriptase reversa associada ao HIV,²⁸ previsão de enantioselectividade associada a ordem de eluição de enantiómeros em HPLC,²⁹ previsão de toxicidade,³⁰ auxílio no diagnóstico de cancro, nomeadamente cancro do pulmão,³¹ previsão de pontos de fusão de líquidos iónicos,⁸ previsão de pKa de compostos monopróticos,³² filtro na separação entre fármacos e não fármacos,³³ estratégia de teste integrada na avaliação de toxicidade no âmbito da legislação REACH,³⁴ identificação estrutural de péptidos tendo como base diferentes padrões de fragmentação MS/MS,³⁵ classificação entre inibidores e não inibidores do factor Xa,³⁶ identificação de inibidores de interações

inter-proteicas,³⁷ previsão de permeabilidade de fármacos no sistema nervoso central,³⁸ classificação de antagonistas da dopamina e serotonina³⁹ e identificação de qual a enzima P450 dominante na metabolização de determinado composto.⁴⁰

No algoritmo usado neste trabalho há uma partição sequencial de objectos em dois subgrupos mais pequenos, destes noutros dois subgrupos mais pequenos cada, e assim sucessivamente, obtendo-se no final uma árvore de decisão (**Figura I.7**) cujos nós terminais são mais homogêneos do que os nós que lhes deram origem relativamente à propriedade em estudo. Grupos que dão origem a subgrupos mais pequenos denominam-se nós parentais (*parental nodes*), caso contrário, denominam-se nós terminais. Os dois nós que resultam de um nó parental denominam-se nós descendentes (*child nodes*). Cada nó é produzido por uma regra lógica, definida normalmente por um descritor, em que objectos tendo valor acima de determinado patamar caem num de dois nós descendentes, e objectos possuindo valor abaixo caem no outro nó descendente. A partição pode ser categórica ou numérica; no primeiro caso os objectos vão ser agrupados de uma forma qualitativa (por classes), sendo as árvores assim definidas denominadas árvores de classificação; no segundo caso os objectos vão ser agrupados de uma forma quantitativa sendo o valor atribuído a cada grupo uma média do valor associado a cada objecto membro desse mesmo grupo; as árvores assim definidas denominam-se árvores de regressão (**Figura I.7**).

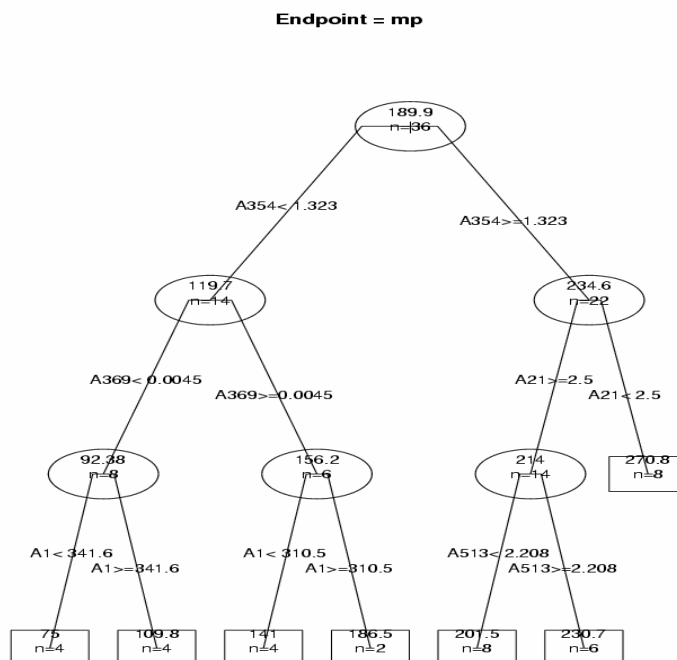


Figura I.7: Representação esquemática de uma árvore de decisão na resolução de um problema de regressão. Grupos que dão origem a outros grupos denominam-se nós parentais (ovais), caso contrário denominam-se nós terminais (rectângulos).

O processo completo de construção da árvore de decisão compreende três passos fundamentais. No primeiro, uma árvore completa é construída através da partição em nós progressivamente mais pequenos, cada nova partição sendo avaliada por uma *função de impureza*, função essa que decresce sempre e quando a nova divisão permita que o conteúdo dos nós descendentes seja mais homogêneo relativamente à propriedade em questão. A avaliação do acréscimo de homogeneidade associada a uma partição é avaliada comparando a função de impureza associada ao nó parental $I(t)$ com a mesma função associada a cada um dos nós descendentes ($I(t_L)$ e $I(t_R)$), sendo obtida uma variação da impureza associada a determinada partição $\Delta I(v, t)$ para uma determinada variável v (descritor) e determinado nó t :

$$\Delta I(v, t) = I(t) - p_L I(t_L) - p_R I(t_R)$$

Equação I.5

onde p_L e p_R são as proporções de objectos provenientes do nó parental t que caíram no nó descendente direito e esquerdo respectivamente.

Existem muitas funções de impureza definidas, mas para árvores de regressão a função de impureza mais utilizada é a seguinte:

$$I(t) = \sum_{x_n \in t} (y_n - \bar{y}(t))^2$$

Equação I.6

em que t está associado a um determinado nó, y_n representa o valor experimental da propriedade em questão associada a um objecto x_n e $\bar{y}(t)$ é a média ou mediana do valor associado aos objectos constituintes de determinado nó.

No caso das árvores de classificação utilizadas neste trabalho a função de impureza utilizada foi o índice de impureza *Gini* definido por:

$$I(t) = 1 - \sum_{i=1}^k (p_i(t))^2$$

Equação I.7

em que $p_i(t)$ é a proporção de objectos da classe i no nó t , sendo assim contabilizadas todas as k classes em estudo. Desta forma o parâmetro de impureza dá-nos o erro associado a cada nó.

No segundo passo um conjunto de árvores mais pequenas (sub-árvores) é obtido através da obliteração (*pruning process*) de determinados nós da árvore obtida no primeiro passo. A selecção dos nós mais fracos é baseada na medida de custo complexidade a qual irá decidir qual das sub-árvores, contendo o mesmo número de nós terminais irá permitir a obtenção de um erro mais baixo.

$$R_{\alpha}(t) = R(t) + \alpha \left| \tilde{T} \right|$$

Equação I.8

Em que $R_{\alpha}(T)$ é a medida de custo complexidade, $R(T)$ é o erro associado aos nós de determinada sub-árvore. No caso da regressão o erro baseia-se na soma dos quadrados das diferenças entre valores experimentais e valores previstos pelo modelo. Para a classificação o erro baseia-se no número de objectos mal classificados. A segunda parcela da **equação I.8** é uma componente de penalização proporcional ao tamanho da árvore em que é o número de nós terminais na sub-árvore e α é um parâmetro de complexidade, associado à profundidade da árvore.

Finalmente no terceiro passo, partindo do conjunto de árvores seleccionadas (árvores mais pequenas são sub-árvores das maiores) com o valor mais baixo de custo-complexidade para determinado número de nós terminais, vai ser seleccionada a árvore que irá permitir obter um erro mais baixo em validação cruzada (ver glossário).

I-1.3: *Random Forest*

Random Forest é um método de aprendizagem automático desenvolvido por Leo Breiman aparecendo descrito em 2001⁴¹ num artigo com mais de 500 citações! Este método consiste na construção de um conjunto de árvores de decisão (**Figura I.8**) tal como as descritas no tópico anterior, a única diferença é que as árvores construídas para as *Random Forest* não passaram pelo processo de obliteração (*pruning*) nem pelo processo de selecção por validação cruzada.

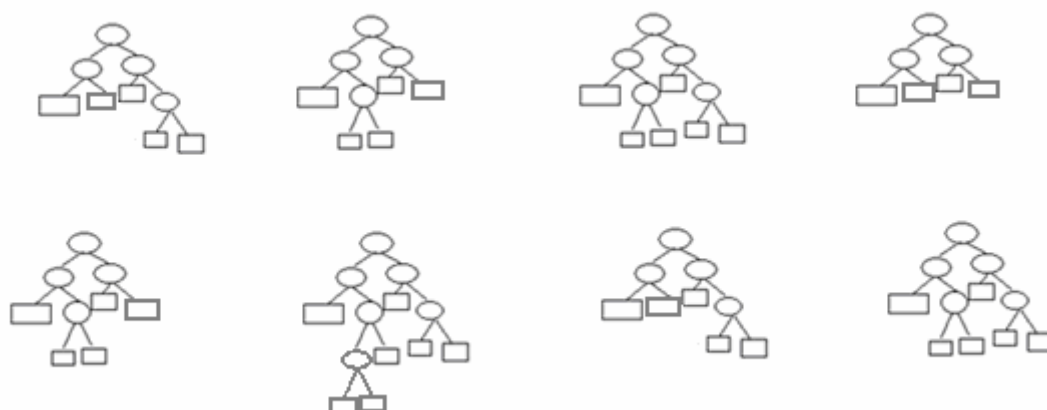


Figura I.8: Conjunto de árvores de decisão, constituindo uma *Random Forest*.

No final a previsão final da *Random Forest* é a previsão maioritária (na classificação) ou a média (na regressão) entre o conjunto de árvores de decisão construídas. A variabilidade entre as árvores construídas está assegurada já que cada árvore é obtida a partir de uma fracção do conjunto de treino escolhida aleatoriamente, normalmente uma fracção de $1-e^{-1}$, ou seja *ca* de 2/3 dos objectos originais são usados (amostra *bootstrap*). Os objectos em falta são substituídos por repetições de objectos permanentes, ficando *ca* de 1/3 dos objectos originais de fora (amostras *Out of Bag*). O segundo factor que contribui para a variabilidade entre as árvores baseia-se no facto de a escolha do descritor que produz a melhor divisão para um dado nó estar restringida a um número *mtry* de descritores seleccionados aleatoriamente e especificamente para aquela partição.⁴² Este método, ao contrário das redes neuronais, funciona bem com grandes conjuntos de descritores, não havendo necessidade de selecção prévia dos mesmos para que a capacidade de previsão seja elevada, muito embora recentemente tenha ficado provado que é possível melhorar substancialmente as capacidades de previsão de uma *Random Forest* através de uma escolha judiciosa de descritores moleculares tendo ficado descrito um procedimento para essa selecção.⁴³ Para além do modelo funcionar bem com um conjunto elevado de descritores, as *Random Forest* podem constituir modelos com elevada capacidade de previsão mesmo não havendo um grande investimento na afinação dos parâmetros que controlam a construção da *Random Forest*, essencialmente o número de árvores e o número *mtry* de descritores seleccionados aleatoriamente para a obtenção de cada partição.

O algoritmo *Random Forest* oferece a possibilidade de obtenção de um conjunto de medidas úteis para avaliar a capacidade e características do modelo na sua globalidade ou analisar detalhadamente cada objecto submetido ao modelo. Entre as medidas mais úteis destaca-se o valor da capacidade de previsão obtida para a amostra *Out of Bag*. Esta medida baseia-se nas previsões obtidas por cada árvore do modelo para cada parcela do conjunto de treino que não tinha sido usada para a

construir. Esta medida é particularmente robusta por se basear em previsões para objectos que não foram usados para obter as árvores que permitiram a obtenção dessas previsões.

Outra medida relevante proporciona o valor da importância de cada um dos descritores usados para construir o modelo. Para cada descritor os valores associados a cada um dos objectos (quando estão em situação *Out of Bag*) vão ser substituídos por valores aleatórios sendo posteriormente comparada para cada árvore a diferença entre as previsões obtidas por modelo com determinado descritor aleatório e modelo original, sendo posteriormente calculada uma média para todas as árvores. Aqueles descritores que provocaram maiores oscilações entre modelo com o preciso descritor aleatório e modelo original vão ser aqueles descritores mais importantes. No caso da classificação a medida de importância é definida matematicamente por:

$$importância = \frac{\sum_{i=1}^p (f_{i-aleatório} - f_i)}{\frac{p}{\sqrt{p}} \text{desvio_padrão}(f_{aleatório} - f)}$$

Equação I.9

O valor da importância expressa na **Equação I.9** é obtida tendo em conta a totalidade das p árvores usadas na construção do modelo em que $f_{i-aleatório}$ e f_i são as fracções de objectos mal classificados pela árvore i considerando respectivamente modelo com determinado descritor aleatório e modelo sem qualquer descritor aleatório. Por sua vez, $\text{desvio_padrão}(f_{aleatório} - f)$ é o desvio padrão dos desvios entre modelo com descritor aleatório e modelo sem descritor aleatório considerando todas as p árvores.

No caso da regressão vai ser considerada a diferença entre M_j e M considerando respectivamente modelo com descritor j aleatório e modelo sem alteração, em que M ou M_j é obtido considerando todos os p objectos, do quadrado da diferença entre valor previsto pelo conjunto de árvores $Y^{OOB}(X_i)$ e valor experimental para cada objecto Y_i (**Equação I.10**).

$$M_ou_M_j = \frac{1}{p} \sum_{i=1}^p \{Y^{OOB}(X_i) - Y_i\}^2$$

Equação I.10

A terceira medida relevante obtida pela *Random Forest* proporciona um valor de proximidade (ou semelhança) entre um determinado par de objectos. Essa medida de proximidade está definida como a proporção de árvores da *Random Forest* em que os dois objectos coincidiram no mesmo nó terminal.

Pela sua capacidade de previsão e medidas relevantes obtidas, o método das *Random Forest* é bastante atractivo em modelação tendo sido descrito o seu uso numa grande variedade de aplicações tais como a previsão do potencial de um composto em provocar dermatite de contacto alérgica,⁴³ classificação de reacções fotoquímicas,¹⁸ classificação de reacções metabólicas,¹⁹ previsão de mutagenicidade,⁴⁴ previsão de solubilidade aquosa,⁴⁵ classificação de substâncias proibidas usadas como agentes dopantes no desporto,⁴⁶ previsão de ponto de fusão, solubilidade aquosa e coeficiente de partição octanol/água de fármacos,⁴⁷ no varrimento sequencial de grandes bases de dados para a definição de regiões do espaço químico com maior potencial em conter maior número de moléculas activas como fármacos,⁴⁸ obtenção de modelos de regioselectividade de substratos dos citocromos P450: 3A4, 2D6 e 2C9, enzimas importantes na metabolização de fármacos,⁴⁹ classificação de fármacos segundo o grau de metabolização e solubilidade aquosa,⁵⁰ previsão de actividade anti-cancerígena,⁵¹ previsão da qualidade de sobreposição na comparação entre ligandos a associar a proteínas,⁵² estudo QSAR da actividade anti-microbiana de 3-nitrocoumarinas e compostos associados⁵³ e modelação do grau de afinidade entre o canal de potássio codificado pelo gene hERG e determinado fármaco, afinidade essa associada à ocorrência de morte súbita em humanos⁵⁴.

I-1.4: Support Vector Machines

Support Vector Machines (SVM) é um método de aprendizagem automática introduzido e desenvolvido, na sua forma actual, nos laboratórios AT&T Bell por Vapnik e colaboradores durante a década de 90 (1992-1997), muito embora as origens deste método remontem à década de 60, quando Vladimir Vapnik e colaboradores desenvolveram o algoritmo *Generalized Portrait*, que consiste na obtenção de um hiperplano linear que separe eficazmente dois tipos diferentes de objectos.⁵⁵ O método *Support Vector Machines* é uma generalização desse algoritmo incluindo casos não separáveis linearmente, sendo actualmente um método muito popular entre a comunidade científica principalmente pela sua capacidade para resolver problemas complexos.

Entre os muitos estudos publicados que utilizam *Support Vector Machines* é possível destacar alguns tal como o estudo de inibição de enzimas da superfamília citocromo P450,⁵⁶ com papel extremamente importante na metabolização de fármacos, ainda a referir o estudo de previsão relativamente à classe estrutural associada a uma proteína⁵⁷ ou os estudos de previsão da ocorrência de clivagem de oligopéptidos pela protease do HIV-1⁵⁸ ou a classificação relativamente à

carcinogenicidade de compostos N-nitroso,⁵⁹ classificação geográfica de azeites,⁶⁰ estimativa de entalpias de formação,⁶¹ modelação da actividade de quinolonas associada a receptores das benzodiazepinas,⁶² modelação do logaritmo do coeficiente de partição entre octanol e água,⁶³ previsão de diagrama de fases para sais,⁶⁴ selecção de descritores,⁶⁵ previsão de solubilidade aquosa de compostos orgânicos,⁶⁶ estudo de ligação de compostos à trombina, ligação de compostos ao factor Xa, estudo da potência carcinogénica, modelação de energia de dissociação de ligações O-H em fenóis substituídos, mobilidade absoluta em electroforese capilar, modelação da capacidade calorífica, estas e outras aplicações aparecem descritas no artigo de Dudek *et al.*⁶⁷

Este tipo de método engloba dois tipos principais de algoritmos – um para classificação e outro para regressão. Em ambos os casos é optimizado um hiperplano/função que, no caso da classificação, separe pares de classes de objectos ou, no caso da regressão, modele ou explique a propriedade em estudo para os objectos em causa. Assim, ainda que tenham bastantes características em comum, classificação^{68,69} e regressão⁷⁰ baseiam-se em algoritmos substancialmente diferentes, sendo descritos separadamente nos tópicos que se seguem.

I-1.4.1: *Support Vector Machines* para Classificação

Como foi referido anteriormente, o objectivo principal deste algoritmo consiste em encontrar um hiperplano (ou plano multidimensional) que eficazmente separe dois tipos de objectos distintos. A descrição matemática do algoritmo será iniciada para casos linearmente separáveis, adaptando-se posteriormente às funções definidas para a situação mais geral que inclui casos não separáveis linearmente. Começando pela situação mais simples de um caso de duas classes de objectos linearmente separáveis (**Figura I.9**), correspondendo às classes 1 e -1, vai ser encontrada uma condição linear que separe essas duas classes (**Equação I.11**):

$$y(x) = \text{Sinal}(w^T x + b)$$

Equação I.11

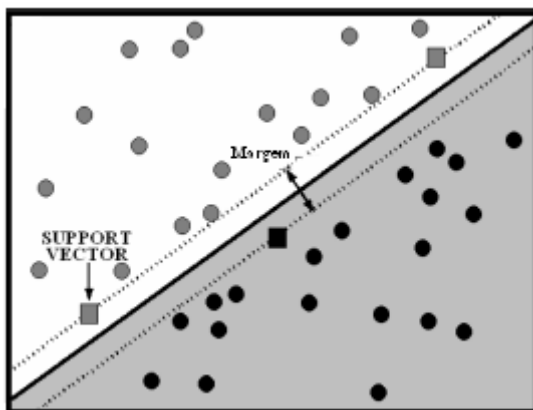


Figura I.9: Exemplo de duas classes linearmente separáveis.⁶⁸

Em que w é o vector transposto dos pesos que o vector das variáveis independentes x vai ter sendo b ordenada na origem, com a função $y(x)$ a dar o sinal (-1 ou 1) que a expressão $w_c^T + b_c$ vai apresentar. A **Equação I.11** baseia-se no hiperplano genérico (w, b) definido para as coordenadas x que respeitem a seguinte condição (**Equação I.12** e representado pela linha a cheio da **Figura I.9**):

$$w^T x + b = 0$$

Equação I.12

que divide o espaço dos objectos em duas regiões de sinais opostos. Assim, um objecto i será correctamente classificado, sem erro associado, se o respectivo vector x_i for projectado numa região do espaço cujo sinal é igual à classe y_i a que pertence, ou pelo menos na fronteira, *i. e.* a seguinte condição tem que ser respeitada para todos os objectos:

$$y_i \cdot (w^T x_i + b) \geq 1$$

Equação I.13

ou seja os objectos têm que estar acima ou abaixo das linhas a tracejado, ou mesmo em cima dessas linhas, (**Figura I.9**) de acordo com a classe a que pertencem. Adicionalmente, pode-se dizer que a separação, num conjunto de objectos correctamente classificado, está optimizada se a distância entre os objectos mais próximos em relação ao hiperplano for maximizada. Na prática, para os objectos mais próximos do hiperplano, a seguinte condição tem que ser respeitada:

$$\min_i \|w^T x_i + b\| = 1$$

Equação I.14

Para estes objectos, a distância em relação ao hiperplano tem que igualar o inverso da norma do vector pesos: $\frac{1}{\|w\|}$.

Assim, o hiperplano que optimiza a separação entre objectos de duas classes é aquele que minimiza w . Assim sendo o processo de optimização pode ser expresso pela seguinte função (**Equação I.15**) que é usada para minimizar w tendo em conta a restrição da **Equação I.13** aplicada exclusivamente aos objectos mais próximos do hiperplano, ou seja *support vectors*.

$$\varphi(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i (y_i [w^T x_i + b] - 1)$$

Equação I.15

Neste processo φ vai ser optimizado relativamente a w e b e posteriormente relativamente a α_i . A optimização da função φ relativamente a w e a b é dada igualando a zero as derivadas em ordem a b e a w . As derivadas assim obtidas vão ser substituídas na **Equação I.15** sendo obtida a seguinte expressão:

$$L_D(\alpha) = \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i y_i (x_i^T x_j) y_j \alpha_j - \sum_{i=1}^l \alpha_i$$

Equação I.16

em que:

$$w_c = \sum_{i=1}^l \alpha_i y_i x_i$$

Equação I.17

A **Equação I.17** foi obtida igualando a zero a derivada da **Equação I.15** em ordem a w . No passo seguinte, a **Equação I.16** vai ser optimizada em relação aos α 's. Olhando um pouco mais detalhadamente, o processo de optimização em relação aos α 's pode ser representada da seguinte forma:

$$\left[\frac{1}{2} \alpha' H \alpha + z' \alpha \right]$$

Equação I.18

em que H tem elementos $h_{ij} = y_i (x_i^T x_j) y_j$ sendo z' um vector de -1's tal como pode ser visto pela **Equação I.16**. A **Equação I.18** é uma equação quadrática, que vai ser resolvida na determinação dos

valores de α . Desta forma, sabendo os valores de α , substituindo a **Equação I.17** na **Equação I.11** vai ser obtida a seguinte expressão classificadora:

$$y(x) = \text{Sinal}\left(\sum_{i=1}^l \alpha_i y_i x_i^T x + b\right)$$

Equação I.19

em que x_i^T é o vector transposto respeitante aos *support vectors*, ou seja, os objectos mais próximos do hiperplano e que respeitam a condição da **Equação I.14** (**Figura I.9 - support vectors**), por sua vez, x refere-se a um objecto genérico. Até aqui foi focado o caso de optimização de hiperplano quando um problema de classificação binário é linearmente separável, neste caso coloca-se a questão relativamente àquilo que é necessário fazer para separar dois tipos de objectos não separáveis linearmente. Neste caso, o algoritmo das *Support Vector Machines* adiciona um passo extra relativamente ao processo de optimização descrito para o caso linear. A resolução deste problema não linear passa por obter novo hiperplano definido, não no espaço x de *input* original, mas num novo espaço de maior dimensionalidade onde os objectos são projectados por intermédio de uma função $\Phi(x)$.

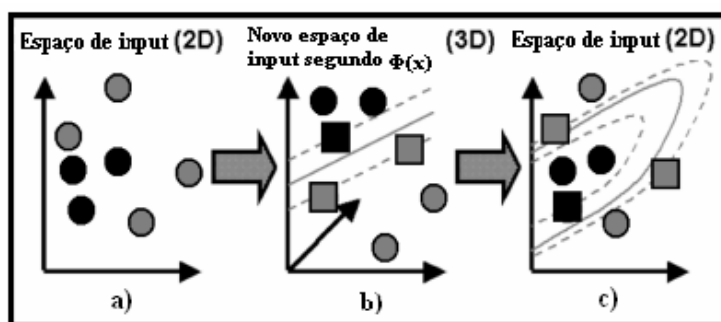


Figura I.10: Caso de problema não separável linearmente no espaço de *input* original a) mas linearmente separável num novo espaço de maior dimensionalidade onde os objectos são projectados segundo $\Phi(x)$ b).⁶⁸ Hiperplano linear em espaço de maior dimensionalidade pode ser convertido em hiperplano não linear c) no espaço de *input* original.

Assim, a **Equação I.16** toma a seguinte forma (**Equação I.20**) substituindo o produto escalar $(x_i^T \cdot x_j)$ por outro produto escalar $\langle \Phi(x_i), \Phi(x_j) \rangle$:

$$L_D(\alpha) = \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i y_i \langle \Phi(x_i), \Phi(x_j) \rangle y_j \alpha_j - \sum_{i=1}^l \alpha_i$$

Equação I.20

Ainda assim a função $\Phi(x)$ não pode ser escolhida livremente. O algoritmo das *Support Vector Machines* restringe-se a um leque restrito de funções que satisfazem a seguinte condição:

$$K(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle$$

Equação I.21

ou seja, o produto escalar entre duas funções que projectam os objectos em causa no novo espaço de *input* ($\langle \Phi(x_i), \Phi(x_j) \rangle$) equivale a uma função no espaço de *input* original $K(x_i, x_j)$. A **Equação I.21** representa um dos atributos mais significativos das *Support Vector Machines*, simplificando significativamente o problema de optimização do hiperplano ao nível de recursos utilizados. Assim o “hiperplano” resultante da optimização pode assumir um padrão mais complexo e deixar de ser um hiperplano linear tal como aparece representado no gráfico *c)* da **Figura I.10**. Matematicamente, substituindo a **Equação I.21** na **Equação I.20** resulta a seguinte expressão:

$$L_D(\alpha) = \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i y_i K(x_i, x_j) y_j \alpha_j - \sum_{i=1}^l \alpha_i$$

Equação I.22

A **Equação I.22**, à semelhança da **Equação I.16**, vai ser optimizada como uma equação quadrática, mas o factor mais relevante, consiste no facto de que, por intermédio da função $K(x_i, x_j)$ em vez do produto escalar $\langle \Phi(x_i), \Phi(x_j) \rangle$, se pode proceder à transformação apresentada na **Figura I.10** passando de uma situação como a representada no gráfico *a)* para outra *c)* projectada no espaço de *input* original sem passar pelo ponto intermédio *b)* de projecção num novo espaço de *input*. Analogamente à função de classificação (**Equação I.19**) para o caso linear, vai existir uma equação mais abrangente incluindo casos não lineares (**Equação I.23**):

$$y(x) = \text{Sinal} \left(\sum_{i=1}^{N_{SVs}} \alpha_i y_i K(x_i, x) + b \right)$$

Equação I.23

Estando a **Equação I.23** estritamente dependente dos *support vectors* tal como aparece expresso. Somente algumas funções *Kernel* $K(x_i, x_j)$ podem ser utilizadas. A função utilizada ao longo deste trabalho foi a função radial:

$$K(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right)$$

Equação I.24 – Função Radial

A grande vantagem da função representada na **Equação I.24** é que só tem um parâmetro σ a otimizar.

De uma forma intuitiva e, devido ao facto de ser possível uma fronteira bastante complexa, existe o risco de *overfitting*. Para evitar essa situação, as *Support Vector Machines* usam um parâmetro adicional de forma a regular a complexidade da fronteira “hiperplano” (**Figura I.11**).

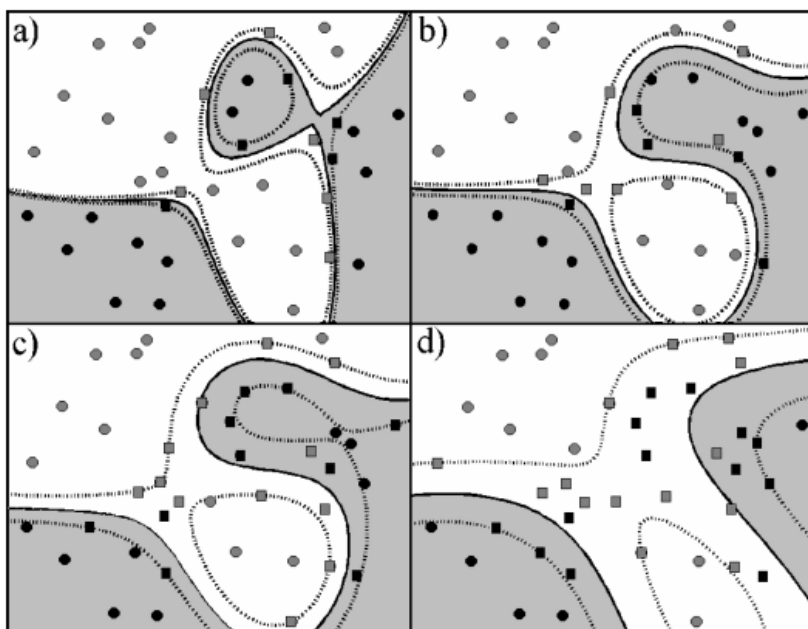


Figura I.11: Variação do grau de complexidade da fronteira entre dois tipos de objectos, desde o caso mais complexo representado em a) até ao caso mais simples d).⁶⁸

Para introduzir este parâmetro adicional há que recordar que o objectivo deste algoritmo é encontrar uma fronteira que separe de uma forma mais acertada e eficaz dois tipos de objectos diferentes, na impossibilidade destes serem perfeitamente separados podem ser autorizados desvios $\xi_i > 0$ para objectos individuais x_i . Os desvios ξ_i representam a distância que um objecto mal classificado apresenta em relação à margem correspondente à classe correcta. Matematicamente o processo de optimização vai requerer, para além da minimização do parâmetro $\frac{1}{2}\|w\|^2$, a minimização do erro empírico, dado pela soma dos desvios autorizados $\sum_{i=1}^l \xi_i$. Tudo isto combinado na seguinte expressão (**Equação I.25**):

$$\varphi(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i$$

Equação I.25

em que C representa o parâmetro da complexidade, ou seja, a importância que vai ser dada aos desvios ξ_i em relação ao limite da classe correcta. Na prática valores de C muito elevados vão corresponder a fronteiras muito ajustadas aos dados experimentais (*a*) **Figura I.11** , ou seja, só pequenos desvios serão tolerados, por sua vez valores baixos de C permitirão tolerâncias maiores aos desvios ficando fronteiras definidas como as do caso *d*) da **Figura I.11**. Seguindo o mesmo raciocínio utilizado para o caso linear vai ser obtida uma função de optimização semelhante à **Equação I.15** e que vai incorporar os parâmetros C e ξ_i representados na **Equação I.25**

Da mesma forma que a **Equação I.15** foi optimizada em ordem a w , b e α , a optimização vai agora ser feita em ordem a w , b , ξ_i , α e α' .

Dado que as *Support Vector Machines* foram desenvolvidas para separar problemas de duas classes, fica a questão de como procederá em situações onde intervêm 3 ou mais classes. Uma das possíveis estratégias, correspondendo àquilo que foi utilizado neste trabalho, consiste em desenvolver diversos modelos *Support Vector Machines*, um para cada combinação de pares de classes e um novo objecto a ser classificado vai ser sujeito a todos esses modelos, no final, a classe mais escolhida será a classe final desse objecto obtida pelo modelo *svm* global, que inclui todos os modelos binários correspondentes a cada combinação de duas classes.

I-1.4.2: Support Vector Machines para Regressão

O conceito deste algoritmo é um pouco diferente do caso de classificação já que, em vez de ser optimizado um hiperplano ou uma fronteira que mais eficazmente separe dois tipos de objectos, vai ser optimizada uma função que modele os valores da propriedade em estudo relativamente aos objectos do conjunto de treino. Essa função, na sua forma mais básica, produzirá desvios até ao patamar máximo ε para cada um dos objectos usados no treino em relação à propriedade em estudo e terá o declive w minimizado. Neste caso, à semelhança da estratégia usada para o algoritmo de classificação, será focado em primeiro lugar o caso linear passando posteriormente para o caso mais geral que inclui problemas não lineares. Assim, a função linear a encontrar pelas *Support Vector Machines* vai ter a seguinte forma genérica:

$$f(x) = \langle w, x \rangle + b$$

Equação I.26

em que $\langle w, x \rangle$ é o produto escalar entre o vector w dos pesos e os parâmetros de *input* de um objecto genérico. De forma a otimizar $f(x)$ o declive w terá que ser minimizado, o que na pratica corresponde a minimizar $\frac{1}{2}\|w\|^2$. Adicionalmente terão que se cumprir as seguintes condições para que só desvios iguais ou inferiores a ε sejam tolerados:

$$\begin{aligned} y_i - \langle w, x_i \rangle - b &\leq \varepsilon \\ \langle w, x_i \rangle + b - y_i &\leq \varepsilon \end{aligned}$$

Equação I.27

Na prática essa precisão ε não é estritamente respeitada e um tipo de desvio adicional ξ é tolerado (**Figura I.12 a**), semelhante ao desvio introduzido a partir da **Equação I.25** para o caso da classificação

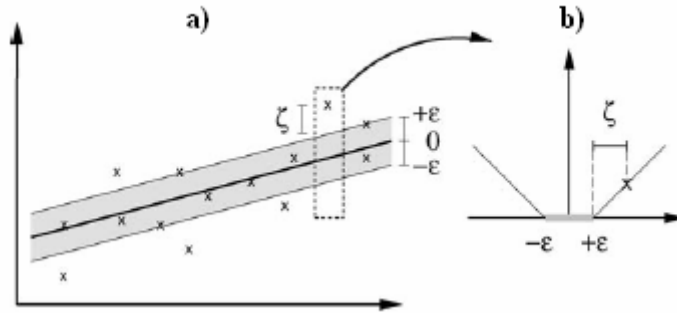


Figura I.12: Exemplo de função linear onde desvios superiores a ε são tolerados a) sendo ξ o excesso em relação a ε . À direita b) aparece um tipo de função penalizadora que lida com desvios superiores a ε .⁷⁰

aparecendo definidas ξ_i e ξ_i^* para casos de desvios positivos e negativos respectivamente (**Figura I.12**). Introduzindo este conceito do desvio ξ adicional de uma forma matemática, terá que ser realizado o seguinte procedimento de optimização:

$$\text{Minimizar} \left(\frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i + \xi_i^* \right)$$

Equação I.28

Adicionalmente vai ser definida uma função penalizadora que lida com desvios superiores a ε (**Figura I.12 b**) definida matematicamente como:

$$|\xi|_{\varepsilon} = \begin{cases} 0 & \text{se } |\xi| \leq \varepsilon \\ |\xi| - \varepsilon & \text{restantes casos} \end{cases}$$

Equação I.29

em que $|\xi|$ equivale a $|f(x) - y|$. Em teoria existem outras funções penalizadores mas a representada pela **Equação I.29** (**Figura I.12 b**) é a mais implementada. Incorporando a informação da **Equação I.29** na **Equação I.27** são obtidas as seguintes condições (**Equação I.30**) constituindo restrições adicionais ao problema de minimização da **Equação I.28**:

$$\begin{aligned} y_i - \langle w, x_i \rangle - b &\leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i &\leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* &\geq 0 \end{aligned}$$

Equação I.30

Desta forma, tendo em conta a **Equação I.28** e a **Equação I.30** a seguinte expressão resultante será otimizada:

$$\begin{aligned} L_D = & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i + \xi_i^* - \sum_{i=1}^l (\eta_i \xi_i + \eta_i^* \xi_i^*) \\ & - \sum_{i=1}^l \alpha_i (\varepsilon + \xi_i - y_i + \langle w, x_i \rangle + b) \\ & - \sum_{i=1}^l \alpha_i^* (\varepsilon + \xi_i^* + y_i - \langle w, x_i \rangle - b) \end{aligned}$$

Equação I.31

A **Equação I.31** é semelhante à **Equações I.15**, sendo esta usada no problema de classificação. Relativamente à optimização esta vai processar-se de forma semelhante à optimização respeitante ao problema de classificação. De uma forma semelhante ao caso da classificação, a **Equação I.31** vai ser optimizada em relação a w, b, ξ_i, ξ_i^* igualando a zero as derivadas parciais em relação a cada uma destas variáveis, as expressões resultantes vão ser substituídas na **Equação I.31** resultando a seguinte expressão:

$$L_D = -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*)$$

Equação I.32

Posteriormente, a **Equação I.32** vai ser otimizada em relação α_i e α_i^* descobrindo os zeros da **Equação I.32**, à semelhança do caso de classificação, sendo α_i e α_i^* as variáveis independentes desta Equação. Tendo em conta a derivada parcial da **Equação I.31** em ordem a w , a seguinte igualdade é obtida:

$$w = \sum_{i=1}^l (\alpha_i - \alpha_i^*) x_i$$

Equação I.33

que substituindo na **Equação I.26** irá resultar a seguinte expressão:

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b$$

Equação I.34

Só objectos que respeitem a condição $|f(x_i) - y_i| \geq \varepsilon$ (**Figura I.12 a**) é que vão ser utilizados neste processo de optimização sendo considerados *support vectors*. Neste caso o conceito de *support vector* é substancialmente diferente do caso de classificação.

Até agora foi explicada a optimização de uma função linear que modele a propriedade em causa. Numa situação mais geral, terá que ser encontrado um procedimento válido também para problemas não lineares. A ideia original consistia em encontrar uma função linear num outro espaço de maior dimensionalidade onde cada nova dimensão seria função do espaço de *input* original, e a função optimizada em vez de estar representada pela **Equação I.34** seria representada por (**Equação I.35**):

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \langle \Phi(x_i), \Phi(x) \rangle + b$$

Equação I.35

Na prática o produto escalar $\langle \Phi(x_i), \Phi(x) \rangle$ está restringido a uma série de funções Kernel $K(x_i, x)$, tal como foi referido para o caso da classificação (ver **Equação I.21**). Tendo em conta a **Equação I.21**, a **Equação I.35** apresenta-se da seguinte forma:

$$L_D = -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(x_i, x_j) - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*)$$

Equação I.36

A função Kernel utilizada neste trabalho é a função Kernel radial (**Equação I.24**). Por sua vez a função $f(x)$ resultante da optimização vai ter a seguinte representação genérica:

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x_j) + b$$

Equação I.37

No processo de optimização, pode-se considerar a optimização do parâmetro de precisão do modelo ε , havendo uma relação linear entre o nível de ruído do modelo e o valor de ε ,⁷⁰ existindo uma forma de otimizar este parâmetro. Na prática a **Equação I.28** base na optimização da expressão modeladora assume a seguinte forma (**Equação I.38**):

$$\text{Minimizar} \left(\frac{1}{2} \|w\|^2 + C \sum_{i=1}^l c(\xi_i) + c(\xi_i) - l v \varepsilon \right)$$

Equação I.38

À semelhança da **Equação I.28**, esta expressão está sujeita às restrições da **Equação I.30**, seguindo-se um raciocínio idêntico ao seguido até aqui para obtenção da **Equação I.37** optimizada. Fundamental nesta nova optimização é o factor v que é utilizado na obtenção do parâmetro ε óptimo.

O processo de optimização é praticamente idêntico ao algoritmo de classificação descrito no tópico anterior, tendo sido procurado fazer a ponte entre algoritmo de classificação e regressão, ainda assim e devido à natureza específica dos problemas a resolver por cada um destes dois métodos, os conceitos utilizados são diferentes, por um lado a fronteira no caso da classificação, por outro o conceito de função modeladora no caso da regressão, resultando em conceitos diferentes de *support vector* para o caso destes dois tipos de problema.

I-2: Descritores

Até agora foram focados os algoritmos computacionais usados ao longo deste trabalho de forma a modelar, classificar e mapear objectos multidimensionais. Esses objectos em concreto podem ser reacções, moléculas, sais, ligações químicas ou até mesmo átomos. A codificação desses objectos numa representação numérica é vantajosa para que os algoritmos computacionais possam aprender através de exemplos e assim encontrar relações entre a codificação numérica de cada objecto (*input*) e

a respectiva propriedade em estudo. Seguidamente, vão ser abordados os diversos tipos de códigos numéricos (descritores) usados ao longo deste trabalho.

I-2.1: Descritores Calculados pelo Programa Dragon

Trata-se de um conjunto de descritores moleculares desenvolvidos por variados autores e reunidos pelo grupo de R. Todeschini no software Dragon, que é hoje comercializado pela empresa italiana TALETE.⁷¹ Estes descritores estão acessíveis através de duas vias, *online*⁷² (1666 descritores moleculares) e através de programa disponível na página *web* da empresa TALETE⁷¹ (3224 descritores moleculares). Ambas as versões geram 5 tipos de descritores, classificados relativamente ao nível de informação estrutural que codificam. Assim podem encontrar-se descritores *0D* que representam características globais de uma determinada molécula tal como o peso molecular ou número de átomos, não estando codificada qualquer tipo de informação relativamente à disposição dos átomos na molécula. Outro tipo de descritores são descritores *1D*, que contêm alguma informação estrutural ainda que de forma fragmentária dando conta da presença de alguns grupos de átomos ou funcionalidades. Os descritores *2D* têm em conta a conectividade entre os diversos átomos da molécula. Associando a conectividade à informação referente às coordenadas tridimensionais de cada átomo na molécula são calculados descritores moleculares *3D*. Finalmente são calculados *outros* descritores muito variados, que vão desde o cálculo de cargas até propriedades físico-químicas. Ao longo deste trabalho foram usadas só algumas famílias de descritores que vão ser descritas em seguida.

Descritores Constitucionais: São descritores do tipo *0D*, que fornecem informação global sobre cada molécula, como seja o peso molecular, número de ligações, número de átomos, ou número de ocorrências de um determinado elemento. Seguidamente serão focados alguns dos descritores usados ao longo do trabalho descrito nesta dissertação:

- a) Estado electrotopológico – Para cada átomo de uma molécula é contabilizado o seu ambiente topológico e electrónico, resultando a seguinte expressão matemática:

$$I_i = \frac{\left(\frac{2}{Q_i}\right)^2 (Z_i^v - H_i) + 1}{\deg_i}$$

Equação I.39

Nesta expressão (**Equação I.39**) Q_i é o número quântico principal associado à camada de valência do átomo i , Z_i^v é o número de electrões de valência, H_i é o número de hidrogénios ligados a esse átomo e \deg_i é o grau de ramificação desse átomo. A partir do

valor de estado electrotopológico associado a cada átomo pode ser calculada uma média para determinada molécula.

- b) Electronegatividade de Sanderson⁷³ – É baseada no princípio segundo o qual o poder de um átomo para atrair electrões está dependente da carga nuclear efectiva sentida pelos electrões de valência. Tendo em conta que a carga nuclear efectiva aumenta quando o tamanho do átomo diminui, Sanderson achou uma razão de proporcionalidade entre a electronegatividade, S , e o grau de compactação associado a determinado átomo. Essa razão aparece representada matematicamente por:

$$S = \frac{D}{D_0}$$

Equação I.40

em que D é a densidade electrónica (número atómico dividido pelo volume atómico, este determinado pela potência de grau 3 do raio covalente) sendo D_0 a densidade electrónica esperada para um determinado elemento, calculada por extrapolação de densidades electrónicas entre elementos de gases nobres. A partir do valor de electronegatividade atómica pode ser calculado um valor global para uma molécula somando as contribuições individuais de cada átomo.

Fragmentos átomo-centrados: São descritores do tipo $1D$ baseados na contagem do número de átomos específicos associados a um ambiente químico específico. Como exemplo pode considerar-se o número de átomos de flúor ligados a carbonos sp^2 ou o número de átomos de hidrogénio ligados a heteroátomo.

Grupos Funcionais: São descritores do tipo $1D$ que codificam a presença de determinadas funcionalidades na molécula, como aminas, ou tioésteres.

Geométricos: São descritores do tipo $3D$ dependentes da conformação de determinada molécula baseados na sua geometria e em distâncias interatómicas. Os dois descritores geométrico usados ao longo deste trabalho são os índices gravitacionais G_1 e G_2 . Descritos matematicamente pelas seguintes expressões.

$$G_1 = \sum_{i=1}^{A-1} \sum_{j=1+1}^A \frac{m_i \cdot m_j}{r_{ij}^2}$$

Equação I.41

$$G_2 = \sum_{b=1}^B \left(\frac{m_i \cdot m_j}{r_{ij}^2} \right)_b$$

Equação I.42

Neste caso m_i e m_j são as massas atômicas de pares de átomos i e j de determinada molécula. Para G_1 são contabilizadas todas as combinações diferentes de pares de átomos, no caso de G_2 são consideradas todas as combinações de pares de átomos ligados entre si. Por sua vez r_{ij} corresponde à distância inter-atômica para cada respectivo par de átomos. O cálculo deste tipo de descritores requer um modelo tridimensional para a molécula.

Propriedades Moleculares: São descritores que representam determinadas propriedades moleculares, umas definidas teoricamente, outras mensuráveis experimentalmente e calculadas de forma empírica. As propriedades moleculares usadas neste trabalho são o índice de insaturação, refratividade molar e o coeficiente de partição octanol-água.

Topológicos: São descritores do tipo $2D$ que codificam tanto a conectividade da molécula como o seu tamanho, forma e ramificação. Foram utilizados alguns descritores topológicos ao longo deste trabalho tais como:

- a) Índices de forma de Kier⁷⁴ – Medidas de forma definidas matematicamente segundo percursos de um número predefinido de ligações seguidas; para percursos de 1, 2 e 3 ligações estão definidas as seguintes expressões matemáticas:

$$^1k = \frac{A(A-1)^2}{P_1^2}$$

Equação I.43

$$^2k = \frac{(A-1)(A-2)^2}{P_2^2}$$

Equação I.44

$$^3k = \begin{cases} \frac{(A-3)(A-2)^2}{P_3^2} \text{ para } A \text{ par } (A > 3) \\ \frac{(A-1)(A-3)^2}{P_3^2} \text{ para } A \text{ ímpar } (A > 3) \end{cases}$$

Equação I.45

Nestas expressões A representa o número de átomos na molécula. P_1 , P_2 e P_3 são o número de percursos de tamanho 1, 2 e 3 respectivamente.

- b) Índices de forma de Randić⁷⁵ – obtidos considerando todos os átomos de uma molécula, através do quociente entre o número de percursos possíveis (tendo em conta sequências de átomos consecutivos) e o número de combinações de determinado número de átomos (pode ser qualquer combinação de átomos mesmo que não sejam consecutivos).
- c) Índice de Ramificação⁷⁶ – codificação numérica do grau de ramificação de determinada molécula aparecendo definido pela seguinte expressão:

$$Ram = \sum d_i - 2$$

$$d_i \geq 3$$

Equação I.46

O somatório da **Equação I.46** é efectuado para todos os átomos i cujo grau de substituição d_i é igual ou superior a 3. Na prática estão a ser contabilizados todos os átomos que contribuam para a ramificação da cadeia linear principal de determinada molécula

- d) Índice de forma 2D de Petitjean⁷⁷ – Outra maneira de codificar a forma de uma molécula, partindo do conceito de centro e extremo de uma molécula e através desses conceitos definir os parâmetros R (raio), que representa a distância entre um centro e um extremo de uma molécula, e D (diâmetro), que representa a distância associada a extremos dessa mesma molécula. Usando estes dois parâmetros define-se este índice por:

$$PJI2 = \frac{D - R}{R}$$

Equação I.47

- e) Índice de Conectividade, Excentricidade (CSI)⁷⁸ – É uma medida que combina os conceitos de excentricidade (E_i) e conectividade (V_i) associado a determinado átomo e aparece definido por:

$$CSI = \sum_{i=1}^n E_i \cdot V_i$$

Equação I.48

O somatório expresso na **Equação I.48** tem em conta todos os n átomos da molécula em causa. O parâmetro E_i representa a distância entre um átomo i e um átomo j o mais afastado possível de i (a distância considerada é a mais curta no caso de haver diferentes percursos a separar os átomos em questão). Por sua vez V_i é o grau de ramificação associado ao átomo i .

- f) Outros índices de Excentricidade⁷⁹ – Tendo em conta a definição de E_i expressa no tópico anterior, a excentricidade molecular ECC, é definida por:

$$ECC = \sum_{i=1}^n E_i$$

Equação I.49

Nesta expressão o somatório tem em conta todos os n átomos da molécula. A partir do valor de ECC podem ser calculados outros dois parâmetros de excentricidade. O valor de excentricidade atómico médio AECC aparece definido matematicamente por:

$$AECC = \frac{ECC}{n}$$

Equação I.50

Por sua vez o valor de DECC aparece definido como o somatório das diferença entre o valor de E_i definido para cada átomo e o valor de AECC, funcionando como uma medida dos desvios de distâncias inter-atômicas máximas em relação à distância máxima média para uma molécula como um todo. Matematicamente a expressão é a seguinte:

$$DECC = \frac{\sum_{i=1}^n (E_i - AECC)}{n}$$

Equação I.51

- g) Índices de Conteúdo de Informação⁸⁰ (Information content indexes) – Medidas de simetria de determinada molécula comparando as vizinhanças dos diversos átomos até uma determinada esfera k . Assim, são calculadas diversas medidas:

$${}^k IC = - \sum_{i=1}^k \frac{n_i}{n} \log_2 \frac{n_i}{n}$$

Equação I.52

$${}^k SIC = \frac{{}^k IC}{\log_2 n}$$

Equação I.53

$${}^k CIC = \log_2 n - {}^k IC$$

Equação I.54

$${}^k BIC = \frac{{}^k IC}{\log_2 q}$$

Equação I.55

Para estas equações n_i representa o número de átomos da mesma classe, n representa o número total de átomos da molécula e q é o número de ligações da molécula. Assim ${}^k IC$ é o índice de conteúdo de informação considerando as imediações de cada um dos átomos da molécula até uma esfera de vizinhança k , ${}^k SIC$ é o índice de conteúdo de informação estrutural, ${}^k CIC$ é o índice de conteúdo de informação complementar e ${}^k BIC$ é o índice de conteúdo de informação de ligação.

I-2.2: Descritores Baseados no Programa PETRA

PETRA⁸¹ é o nome de um programa que inclui diversos métodos empíricos, desenvolvidos no grupo de investigação de J. Gasteiger, para cálculo de propriedades físico-químicas associadas a moléculas, ligações químicas e átomos, sendo quantificados os seguintes efeitos:

- Calores de formação
- Energias de dissociação de ligação
- Distribuição de cargas sigma
- Distribuição de cargas pi
- Cargas totais
- Efeito de ressonância e energias de deslocalização
- Efeito de polarizabilidade

O *output* deste programa pode ser utilizado no cálculo de diversos tipos de descritores moleculares como sejam os descritores RDF (Radial Distribution Function), SURFACE, XPET e MOLMAPS. Estes descritores vão ser explicados em maior pormenor em seguida tal como os métodos empíricos associados ao programa PETRA.

I-2.2.1: PETRA

Como já foi dito anteriormente, este programa inclui uma série de métodos para cálculos de propriedades físico-químicas associados a uma série de efeitos cujos métodos de cálculo serão aqui descritos.

Distribuição de cargas sigma

O conceito de cargas parciais em átomos de uma molécula está muito difundido na comunidade científica. Apesar deste conceito mostrar uma visão muito simplificada da realidade relativa à distribuição electrónica numa molécula, ele tem-se revelado muito útil na explicação de certos fenómenos. O cálculo das cargas atómicas baseia-se no conceito de electronegatividade definido por Mulliken como a média entre o potencial de ionização IP e a afinidade electrónica EA :

$$\chi = \frac{IP + EA}{2}$$

Equação I.56

No método implementado pelo PETRA as diversas orbitais atómicas são consideradas separadamente, cada orbital tendo um valor de electronegatividade χ_i dependente do estado de hibridação e do nível de ocupação n ($n = 0, 1, 2$ correspondendo a cargas de +1, 0 e -1 respectivamente). A electronegatividade associada a uma orbital i de um dado átomo A está relacionada com a carga associada a esse átomo (**Equação I.57**). Tendo em conta essa relação, ambas as propriedades são calculadas de forma iterativa usando o método PEOE (Partial Equalization of Orbital Electronegativity) em três passos sequenciais:

Passo 1 – Para cada orbital i de cada átomo A :

$$\chi_{iA} = a_i + b_i q_A + c_i q_A^2$$

Equação I.57

Passo 2 – Para cada ligação química A-B:

$$\Delta q_{AB}^{(n)} = \frac{1}{\chi_{iA} + \chi_{iB}} (\chi_{iA} - \chi_{iB}) \alpha^n$$

Equação I.58

$$\Delta q_{AB} = \Delta q_{AB} + \Delta q_{AB}^{(n)}$$

Equação I.59

Passo 3 – Para todas as ligações de um átomo:

$$q_A^{(n)} = \sum \Delta q_{AB}^{(n)}$$

Equação I.60

$$q_A = q_A + q_A^{(n)}$$

Equação I.61

Em todo este processo há que considerar α como o factor de *damping* com valor 0,5. Por sua vez n é o número da iteração, se n for inferior ao número máximo de iterações consideradas, volta-se ao passo 1, agora usando as cargas determinadas na iteração anterior, por sua vez os parâmetros a , b e c na **Equação I.57** são determinados através de valores de electronegatividade tabelados para as cargas +1, 0 e -1 associadas a cada orbital. No final são obtidos os seguintes parâmetros: $q_{A,\sigma}$ (carga sigma associada ao átomo A), $\chi_{A,\sigma}$ (electronegatividade sigma associada ao átomo A), $\Delta q_{AB,\sigma}$ (diferença entre cargas sigma de átomos A e B da ligação), $\Delta \chi_{AB,\sigma}$ (diferença entre electronegatividade de átomos A e B de uma ligação), $SQIT_{AB,\sigma}$ (soma de cargas Q trocadas durante as iterações).

Distribuição de cargas π

Para a realização do trabalho descrito nesta tese foram usadas duas versões do PETRA, a versão 3.11 e a versão 4.0., os métodos usados para calcular as cargas π são substancialmente diferentes. O método utilizado pela versão 3.11, denominado PEPE (Partial Equalization of π -

electronegativity) ainda que proporcione uma versão realista sobre a distribuição de carga nas moléculas, carece de fundamento teórico bem estabelecido. Para resolver esse problema foi introduzido na versão 4.0 um novo método de cálculo baseado na teoria das orbitais moleculares de Hückel (HMO), concebida originalmente para hidrocarbonetos insaturados, e que no PETRA foi adaptada⁸² de forma a assimilar o efeito da presença de heteroátomos, os efeitos indutivos e de hiperconjugação. Ambos os métodos, PEPE e HMO modificado, vão ser descritos nos próximos parágrafos.

a) Método PEPE

O processo de cálculo de distribuição de cargas π tem pontos em comum com o processo descrito anteriormente para a determinação de cargas σ , ou seja é um processo iterativo que usa cargas σ para a determinação da distribuição de cargas π . Este método,⁸³ denominado PEPE (Partial Equalization of π -electronegativity) parte do conceito fundamental de quantificar o peso w que cada estrutura de ressonância vai ter na distribuição electrónica da molécula em questão. O peso w de cada estrutura de ressonância vai ser calculado a partir de duas componentes distintas, a componente topológica w_t e a componente electrónica w_e :

$$w = w_t \cdot w_e$$

Equação I.62

A componente topológica w_t tem em conta, na conversão entre duas estruturas de ressonância, a) se haverá redução no número de ligações covalentes, b) se vai haver decréscimo no número de sistemas aromáticos ou c) se haverá separação de cargas (**Figura I.13**).

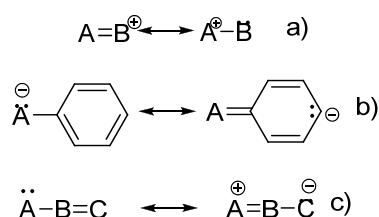


Figura I.13: Tipos de efeitos contabilizados pela componente topológica do peso que determinada estrutura de ressonância vai ter na obtenção de cargas σ . a) alteração no número de ligações covalentes, b) alterações em sistemas aromáticos e c) contabilização da separação de cargas.

Cada um destes três efeitos vai ser representado por um factor, f_a, f_b e f_c respectivamente (**Figura I.13**), sendo usados para calcular w_t :

$$w_t = f_a \cdot f_b \cdot f_c$$

Equação I.63

Relativamente à componente electrónica w_e , o processo de obtenção é substancialmente mais complexo muito embora, à semelhança do cálculo das cargas σ , se baseie no conceito de electronegatividade de cada átomo para definir as trocas de carga entre os átomos intervenientes, definidas num processo iterativo, onde cada estrutura de ressonância vai diferir nas trocas electrónicas processadas entre orbitais π , modeladas pela electronegatividade π (χ_π). Neste processo a electronegatividade associada a cada orbital interveniente numa ligação (σ e π) é definida por:

$$\chi_v = a_v + b_v \cdot q + c_v q^2$$

Equação I.64

em que a_v , b_v e c_v são coeficientes que podem ser ajustados através de dados publicados de potencial de ionização e afinidade electrónica (**Equação I.64**). Por sua vez o valor da carga q na **Equação I.64** é definido por:

$$q = q_\pi + f_{sp} \cdot q_\sigma$$

Equação I.65

Em que a carga q é função das cargas π (q_π) e σ (q_σ), por sua vez f_{sp} é um factor empírico, sendo mais à frente referidas as experiências que permitiram a obtenção deste e outros factores intervenientes neste processo. A carga σ (q_σ) é obtida em cada ciclo através do método descrito anteriormente para determinação de distribuição de cargas σ , por sua vez o valor da carga π é obtida no final de cada ciclo como função das parcelas de carga q_c trocadas em cada ciclo referente ao processo agora em discussão, sendo definido mais adiante a função que define q_c . Segundo as **Equações I.64 e I.65**, vão ser obtidos valores de electronegatividade π (χ_π) e carga q , sendo usados directamente no cálculo do peso w_e :

$$w_e = (\chi_i^\pi - \chi_j^\pi) + f_e \cdot \left(\sum_k^{N_i} q_k - \sum_l^{N_j} q_l \right)$$

Equação I.66

em que a primeira parcela da soma corresponde à diferença de electronegatividades π (χ_π) ou (χ^π) de átomos intervenientes (i e j) na ressonância da molécula e que vêm alterada a sua carga formal nesse processo. Por sua vez, na segunda parcela da soma aparece definida a diferença entre potenciais electrostáticos dos (k e l) átomos vizinhos a cada um dos dois átomos i, j em questão, respectivamente. O potencial electrostático aparece representado na **Equação I.66** como a soma de cargas (q_k ou q_l) de todos os átomos vizinhos k ou l . Por sua vez o factor f_e , à semelhança dos factores anteriores, é obtido empiricamente. Nesta altura, na posse dos valores de w_i e w_e e tendo em conta a **Equação I.62**, pode-se calcular a carga trocada q_c no ciclo actual usando a seguinte equação:

$$q_c = \frac{1}{n_c} \cdot w \cdot f_s$$

Equação I.67

em que n_c corresponde aos ciclos considerados, assim só uma parte da carga q_c é deslocada por cada ciclo, por sua vez f_s corresponde ao sexto e último factor empírico considerado neste processo. Adicionando as mudanças de carga das estruturas de ressonância individuais vai ser obtida a distribuição de cargas π , sendo estas usadas no novo ciclo, na **Equação I.65**. Cada ciclo compreende as **Equações I.62-I.67** para cálculo de cargas π (q_π) sem esquecer as **Equações I.57-I.61** descritas no tópico anterior para determinação das cargas σ (q_σ), que por sua vez vão ser introduzidas na **Equação I.65**. Para finalizar este tópico, importa explicar que os factores $f_a, f_b, f_c, f_e, f_{sp}, f_s$ foram obtidos usando propriedades que dependem directamente da distribuição de cargas, como os desvios químicos de ^{13}C -RMN e sinais obtidos por técnica espectroscópica ESCA. Neste cálculo das cargas π são usadas como base as cargas formais associadas à molécula alvo.

b) Método HMO modificado

O método HMO, inicialmente desenvolvido para hidrocarbonetos insaturados, apresenta como grandes vantagens o facto de assentar no fundamento teórico da mecânica quântica, capaz de explicar muitos fenómenos físico-químicos, e ao mesmo tempo ser um método suficientemente simples para se poder interpretar e poder implementar computacionalmente sem usar grandes recursos. Este método será primeiramente abordado na sua forma original, ou seja, para hidrocarbonetos insaturados e contabilizando em exclusivo sistemas π .

No método HMO, à semelhança de outros cálculos que recorram à mecânica quântica, soluções aproximadas da equação de Schrödinger têm que ser encontradas:

$$H\Psi = E\Psi$$

Equação I.68

em que Ψ é a função de onda, H o operador Hamiltoniano e E a energia total do sistema. Na resolução desta equação, para o problema em questão da determinação da distribuição de cargas π , apenas os electrões π são considerados. A função de onda representando a solução para a **Equação I.68** é uma combinação linear das orbitais (LCAO) $2p_z$ de carbonos da molécula em questão.

$$\Psi = \sum_r^n c_r \varphi_r$$

Equação I.69

Na **Equação I.69**, φ_r é a equação de onda de uma orbital atómica $2p_z$ enquanto c_r é o peso que a equação de onda atómica vai ter na equação de onda molecular Ψ . Os coeficientes c_r podem ser obtidos aplicando o princípio variacional postulando que qualquer função de onda Ψ tem energia associada ε e que esta não pode ser inferior à mais baixa energia E_0

$$\varepsilon = \frac{\int \Psi H \Psi \partial \tau}{\int \Psi^2 \partial \tau} \geq E_0$$

Equação I.70

Para encontrar a mais baixa energia é necessário minimizar ε relativamente aos coeficientes c_r , o que na prática corresponde a igualar a zero a derivada de ε em relação a cada coeficiente c_r .

$$\frac{\partial \varepsilon}{\partial c_r} = 0$$

Equação I.71

Substituindo a **Equação I.69** na **Equação I.70** vai ser obtida a seguinte expressão:

$$\begin{aligned} \varepsilon &= \frac{\int \left(\sum_r c_r \varphi_r \right) H \left(\sum_r c_r \varphi_r \right) \partial \tau}{\int \left(\sum_r c_r \varphi_r \right)^2 \partial \tau} \\ &= \frac{\sum_r \sum_s c_r c_s \int \varphi_r H \varphi_s \partial \tau}{\sum_r \sum_s c_r c_s \int \varphi_r \varphi_s \partial \tau} \end{aligned}$$

Equação I.72

Ψ definida na **Equação I.69**. Este sistema de equações pode ser transposto para uma matriz de Hückel (considerando em exclusivo sistemas π).

$$\begin{bmatrix} \alpha - \varepsilon & \beta_{12} & \dots & \beta_{1n} \\ \beta_{21} & \alpha - \varepsilon & \dots & \beta_{2n} \\ \dots & \dots & \dots & \dots \\ \beta_{n1} & \beta_{n2} & \dots & \alpha - \varepsilon \end{bmatrix}$$

Equação I.76

Para resolver o sistema de equações representado na matriz da **Equação I.76** é necessário fazer simplificações adicionais de forma a aceder facilmente ao valor dos elementos da matriz. O valor de α , ou seja, o valor energético de um electrão numa orbital C-2p_z é posicionado na origem da escala energética da molécula, ou seja, é-lhe atribuído o valor 0. Relativamente aos integrais de ressonância β entre orbitais C-2p_z de átomos que não partilhem uma ligação σ vai ser atribuído valor 0, para os outros casos é assumido o valor $\beta = 1$. Por outro lado ε pode ser obtido como uma função de α e β através do cálculo do determinante da matriz expressa na **Equação I.76**. Desta forma está-se em condições de resolver o sistema de equações e obter os valores dos coeficientes das orbitais atómicas necessários para determinar a distribuição de cargas π ; esta distribuição é obtida realizando uma análise populacional, *i.e.*, somando as fracções de electrões associadas às contribuições que uma determinada orbital atómica vai ter na população das diversas orbitais moleculares. A carga π associada a um átomo r é obtida pela seguinte expressão:

$$q_{\pi,r} = v_r - \sum_{j=1}^{N_{ocupadas}} v_j c_{jr}^2$$

Equação I.77

Com j a percorrer todas as orbitais moleculares ocupadas, e v_r a representar a ocupação da orbital atómica do átomo r , por sua vez v_j representa a ocupação da orbital molecular j sendo c_{jr} o coeficiente da contribuição da orbital atómica r na orbital molecular j .

Até agora foi descrito o método de cálculo de cargas π através da teoria HMO clássica, que não tinha em conta a presença de heteroátomos, para a contabilização destes átomos terá que se ter acesso a novos parâmetros α e β que os representem. Como as energias nestes cálculos têm valores relativos associados estas podem ser expressas em unidades β_{CC} (integrais de ressonância associados a ligação carbono-carbono). Desta forma as seguintes expressões podem ser aplicadas a heteroátomos X :

$$\alpha_X = \alpha_{C-2p_z} + h_X \beta_{CC}$$

$$\beta_{XY} = k_{XY} \beta_{CC}$$

Equação I.78

Com as simplificações, $\alpha_{C-2pz} = 0$ e $\beta_{CC} = 1$, são obtidas as seguintes igualdades:

$$\alpha_X = h_X$$

$$\beta_{XY} = k_{XY}$$

Equação I.79

Os integrais de Coulomb e de ressonância para heteroátomos não podem ser obtidos de um ponto de vista teórico, estes têm que ser determinados através da modelação de uma propriedade externa obtida experimentalmente, sejam sinais obtidos por técnica espectroscópica ESCA, sejam desvios químicos obtidos por ressonância magnética nuclear, sejam valores experimentais de momentos dipolares, tudo propriedades que dependem da distribuição de cargas numa molécula. Existem já disponíveis tabelas com valores de integrais de Coulomb e de Ressonância para uma série de átomos que não o carbono.

A contabilização de heteroátomos presentes numa molécula veio alargar o âmbito de aplicação da teoria de Hückel. Ainda assim há factores muito relevantes que ainda não foram contabilizados, como é o caso das interacções $\sigma-\pi$ – por exemplo, a presença de átomos electroatractores na vizinhança de um sistema π pode induzir a presença de uma carga positiva parcial num átomo de sistema π com essa vizinhança. Esta carga positiva parcial pode resultar numa atracção Coulombica mais acentuada de electrões do sistema π requerendo uma adaptação do integral de Coulomb. Este efeito foi estudado em sistemas π que tivessem associados na vizinhança átomos ou grupos electroatractores mas que não estivessem em ressonância com o sistema π de forma a isolar o efeito de interacções $\sigma-\pi$. Este problema de influência de efeito indutivo na redistribuição de cargas foi tratado por Wheland e Mann, que propuseram uma alteração ao método HMO, sendo este associado a um tratamento iterativo dos integrais de Coulomb, actualizados após cada iteração:

$$\alpha_i = \alpha_0 + \Delta\alpha_i$$

Equação I.80

sendo α_i o integral de Coulomb após cada iteração i , α_0 o integral de Coulomb i inicial, $\Delta\alpha_i$ o valor a adicionar de forma a actualizar o integral de Coulomb naquela iteração. Um dos factores que compõe $\Delta\alpha_i$ é $\Delta\alpha_{\sigma-\pi,k}$ associado a um determinado centro π , k , e é determinado através da média das diferenças entre as electronegatividades residuais $\chi_{\sigma j}$ e a electronegatividade $\chi_{\sigma, \text{metilo}}$, para todos os átomos vizinhos j da primeira esfera topológica em relação ao átomo k em questão. Ainda a referir que $\chi_{\sigma, \text{Cmetilo}}$ é o valor de referência, correspondendo à electronegatividade residual do átomo de carbono do grupo metilo da molécula de etano:

$$\Delta\alpha_{\sigma-\pi,k} = f_{\sigma-\pi} \cdot \frac{1}{N_k} \sum_j^{N_k} (\chi_{\sigma,j} - \chi_{\sigma,Cmetilo})$$

Equação I.81

em que N_k é o número total de átomos vizinhos compreendidos na primeira esfera topológica em relação ao átomo k , por sua vez $f_{\sigma-\pi}$ é um parâmetro empírico. As electronegatividades residuais são obtidas pelo método iterativo PEOE, descrito acima, para a obtenção da distribuição de cargas σ .

De forma a tornar a distribuição de cargas π calculadas o mais próximo da realidade há ainda que ter em conta o efeito de hiperconjugação, habitualmente associado ao efeito doador de grupos alquilo não estando relacionado com efeito indutivo. De forma a contabilizar esse efeito é definido novo parâmetro de correcção $\Delta\alpha_{hip,k}$ para os integrais de Coulomb, definido como:

$$\Delta\alpha_{hip,k} = \sum_j^{N_{alquilo}} f_{hiperconjugacao} \cdot q_{\sigma,j}$$

Equação I.82

em que $q_{\sigma,j}$ é a carga σ associada aos átomos de carbono j , pertencentes a grupos alquilo, adjacentes a átomo de sistema π , k ; por sua vez o parâmetro $f_{hiperconjugacao}$ é um parâmetro empírico. As cargas $q_{\sigma,j}$ são obtidas iterativamente pelo método PEOE. O parâmetro de correcção $\Delta\alpha_{hip,k}$ vai ser incorporado num factor de correcção global $\Delta\alpha_i$ onde se inclui o factor de correcção que contabiliza interacções $\sigma-\pi$ ($\Delta\alpha_{\sigma-\pi,k}$), resultando:

$$\Delta\alpha_i = \Delta\alpha_{\sigma-\pi,k} + \Delta\alpha_{hip,k}$$

Equação I.83

O resultado da **Equação I.83** em cada ciclo iterativo vai ser incorporado na **Equação I.80**.

Cargas totais

As cargas totais são obtidas através da soma entre as cargas σ e π , sendo possível a obtenção de dois parâmetros, $q_{A,tot}$ (carga total do átomo A) e $\Delta q_{AB,tot}$ (diferença entre cargas totais dos átomos A e B de uma ligação química).

Efeito de ressonância e energias de deslocalização

A estabilização de cargas, uma positiva e outra negativa, obtidas através de quebra heterolítica de ligação química podem ser calculadas através da electronegatividade π e da electronegatividade de

electrões não partilhados associadas a cada átomo que esteja em conjugação com os átomos associados à quebra heterolítica de ligação. O processo de encontrar átomos que possam estabilizar uma carga positiva ou negativa não começa com os átomos da ligação que se quebra mas sim com os átomos que têm potencial para estabilizar cargas através de efeitos de ressonância (*source atoms* – átomos origem). Posteriormente, os átomos das ligações que estão conjugadas com os átomos origem vão ser marcados. Podem-se considerar os seguintes tipos de átomos origem:

- átomos doadores (possuem algum par de electrões não partilhado).
- átomos aceitadores (com capacidade de aceitar um par de electrões).
- grupos alquilo associados a ligações múltiplas (efeito de hiperconjugação).
- grupos alquilo associados a átomos doadores (efeito de hiperconjugação).

Estabilização por ressonância é uma propriedade de ligação visto ser uma ligação que se quebra para gerar as cargas. Há duas direcções para quebrar uma ligação de forma heterolítica:

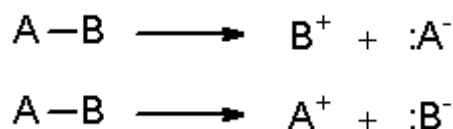


Figura I.14: As duas direcções consideradas na quebra heterolítica de uma ligação química.

Numa ligação química composta pelos átomos A e B convencionou-se que ao primeiro átomo é atribuída carga que está a ser calculada, sendo atribuída carga contrária ao segundo. Um valor R^- é calculado como medida de estabilização de carga negativa, enquanto R^+ é o valor associado à estabilização de carga positiva. Esses valores podem ser combinados num valor adicional R^\pm , que representa a estabilização de cargas em ambos os átomos da ligação. O valor numérico para a estabilização de carga negativa R^- é obtido tendo em conta todos os átomos i que vão estar em ressonância com o centro de carga negativa, neste caso A (**Equação I.84**):

$$R^-(A-B) = \sum_i f^n \chi_{\pi,i}$$

Equação I.84

Ao parâmetro f vai ser atribuído o valor 1 se o átomo da estrutura de ressonância for aromático de outra forma o valor atribuído é 0,67. A variável n representa o número de ligações entre o centro de carga negativa e o átomo i interveniente na deslocalização da carga. O parâmetro $\chi_{\pi,i}$ representa a

electronegatividade π do átomo i . O valor para a estabilização de carga positiva R^+ associado ao átomo A é obtido pela seguinte equação:

$$R^+(A-B) = \frac{c}{\sum f^n \chi_{\pi,i}}$$

Equação I.85

De forma idêntica ao caso anterior, a soma vai-se processar entre todos os átomos i das diversas estruturas de ressonância. Neste caso f e n terão o mesmo significado e aplicação que no caso anterior. O valor de c é uma constante e vai-lhe ser atribuído o valor de 26,63 eV. Por outro lado R^\pm é obtido através da soma entre R^+ e R^- associados a cada um dos dois átomos envolvidos na quebra heterolítica de ligação:

$$R^\pm = R^+(A-B) + R^-(B-A)$$

Equação I.86

No final vão ser obtidas os seguintes parâmetros: R_{AB}^+ – PSTAB –A,B (estabilização por ressonância de carga positiva no átomo A quando ligação A-B sofre quebra heterolítica), R_{AB}^- – NSTAB – A,B (estabilização por ressonância de carga negativa no átomo A quando ligação A-B sofre quebra heterolítica) e R_{AB}^\pm – STABRES –A,B (estabilização por ressonância de carga positiva no átomo A e carga negativa no átomo B).

Tendo ainda em conta a deslocalização de carga, outro tipo de parâmetros são calculados, neste caso medidas directas da deslocalização de carga em vez de medidas de estabilização por ressonâncias. Neste caso os parâmetros calculados são D^+ , D^- , D^\pm . Começando por D^\pm , este parâmetro representa a estabilização mesomérica de carga positiva e negativa nos fragmentos resultantes da quebra heterolítica da ligação em causa (**b – Figura I.15**).

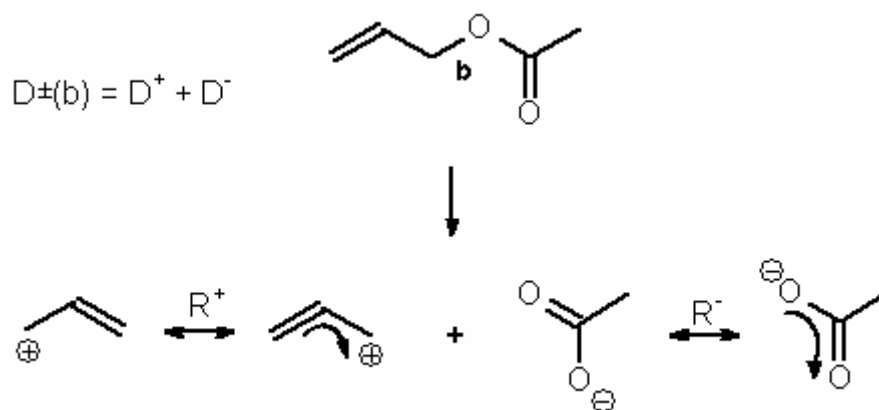


Figura I.15: Quebra heterolítica de ligação b e representação da deslocalização de carga em fragmentos gerados.

É obtido matematicamente pela seguinte soma:

$$D^{\pm}(A \rightarrow B) = D^{+}(A^{+}, A \rightarrow B) + D^{-}(B^{-}, A \rightarrow B)$$

Equação I.87

Neste caso, na quebra heterolítica, a carga positiva vai ser colocada no átomo A e a carga negativa vai estar colocada no átomo B, aparecendo na soma as duas componentes de deslocalização respectivamente. Para cálculo de $D^{+}(A^{+}, A \rightarrow B)$ e $D^{-}(B^{-}, A \rightarrow B)$ vão ser aplicadas as seguintes equações:

$$D^{+}(A^{+}, A \rightarrow B) = \left(\chi_{\pi}(A, q_{\pi}) - \frac{\sum_{i=1}^n \chi_{\pi}(x_i)}{n} \right) \cdot s_{\pi}^{+}$$

Equação I.88

$\chi_{\pi}(A, q_{\pi})$ representa a electronegatividade π associada ao átomo A com carga q_{π} , $\chi_{\pi}(x_i)$ representa a electronegatividade π associada ao átomo x_i interveniente na deslocalização de carga, n representa o número total de átomos intervenientes na deslocalização, por sua vez s_{π}^{+} representa o potencial doador de sistema π em causa, estando representado matematicamente por:

$$s_{\pi}^{+} = \sum_{i=0}^n f_i \cdot m$$

Equação I.89

em que f_i é o potencial doador do átomo ou ligação de tipo específico (valores tabelados em relação a ligação CC olefínica que tem valor convencionado 1, valores calculados pelo método AM1), sendo m o número de ocorrências desse tipo de átomo ou ligação. Por sua vez $D^{-}(B^{-}, A \rightarrow B)$ é calculado de forma semelhante:

$$D^{-}(B^{-}, A \rightarrow B) = - \left(\chi_{\pi}(B, q_{\pi}) - \frac{\sum_{i=1}^n \chi_{\pi}(x_i)}{n} \right) \cdot s_{\pi}^{-}$$

Equação I.90

Nesta equação $\chi_{\pi}(B, q_{\pi})$ representa a electronegatividade π associada ao átomo B com carga q_{π} , $\chi_{\pi}(x_i)$ representa a electronegatividade π associada ao átomo x_i interveniente na deslocalização de carga, n representa o número total de átomos intervenientes na deslocalização, por sua vez s_{π}^{-} representa o potencial aceitador de sistema π em causa, estando representado matematicamente por:

$$s_{\pi}^{-} = \sum_{i=0}^n g_j \cdot k$$

Equação I.91

Nesta expressão, g_j representa o potencial aceitador de determinado tipo de átomo ou ligação j (determinado por método AM1 e relativizado em relação a ligação CC olefínica), em que k é o número de ocorrências do átomo ou fragmento considerado. No final os três tipos de parâmetros calculados vão corresponder a D_{AB}^{+} (estabilização por deslocalização de carga positiva no átomo A quando a ligação AB é quebrada de forma heterolítica), D_{AB}^{-} (estabilização por deslocalização de carga negativa no átomo A quando a ligação A-B é quebrada de forma heterolítica) e D_{AB}^{\pm} (estabilização por deslocalização de carga positiva no átomo A e carga negativa no átomo B).

Efeito de polarizabilidade

Os valores de polarizabilidade considerados no PETRA são a polarizabilidade molecular média associada a uma molécula α_{mol} , a polarizabilidade atômica efectiva associada a um átomo α_{id} , e a polarizabilidade α_b associada a uma ligação química. Todos esses parâmetros são calculados a partir da adição de incrementos associados a cada átomo tendo em conta a sua hibridação específica Os

incrementos estão tabelados e podem ser considerados polarizabilidades estáticas. Assim a polarizabilidade molecular média α_{mol} aparece definida matematicamente como:

$$\alpha_{mol} = \frac{\sum_i^n \alpha_i}{n}$$

Equação I.92

em que são consideradas as polarizabilidades estáticas α_i associadas a cada um dos n átomos da molécula. A outra propriedade considerada é a polarizabilidade efectiva α_{jd} associada a um átomo, que é uma medida de estabilização de carga situada num átomo específico j da molécula, e que é influenciada por todos os átomos da molécula, com peso decrescente com a distância:

$$\alpha_{jd} = \sum_i^n 0,5^{n_{ij}} \cdot \alpha_i$$

Equação I.93

em que α_i é a polarizabilidade estática associada ao átomo i , n_{ij} representa o número de ligações entre o átomo i e o átomo j contendo a carga (**Figura I.16**).

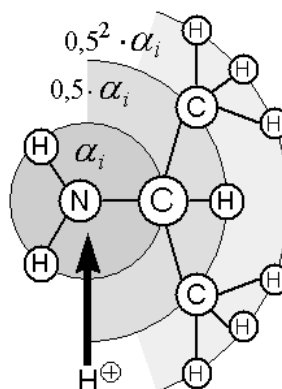


Figura I.16: Estabilização de carga por parte de um átomo específico (neste caso N) influenciada por átomos vizinhos em função da sua proximidade ao átomo em causa. Essa medida de estabilização denomina-se polarizabilidade efectiva.

Por sua vez a polarizabilidade de ligação α_b é calculada como a polarizabilidade média dos dois átomos da ligação.

I-2.2.2: Função de Distribuição Radial - RDF

Algumas das propriedades produzidas pelo PETRA são usadas para calcular descritores moleculares. Os descritores RDF⁸⁴ são um tipo de descritores moleculares que incorporam propriedades atômicas provenientes do PETRA. Os descritores RDF surgiram devido à necessidade de representar estruturas moleculares tridimensionais de tamanhos muito diversos através de códigos numéricos de tamanho fixo. A função de distribuição radial é definida por:

$$g(R) = f \cdot \sum_{i=1}^{A-1} \sum_{j=i+1}^A w_i \cdot w_j \cdot \exp(-\beta(R - r_{ij})^2)$$

Equação I.94

Esta função representa a distribuição de distâncias interatômicas dentro da estrutura molecular. Cada descritor molecular saído desta função corresponde a uma distância R , e se definirmos a mesma gama de R para todas as moléculas do estudo, obtemos o mesmo número de descritores para as moléculas consideradas. Os parâmetros w_i e w_j correspondem a propriedades atômicas associadas aos átomos i e j ; o parâmetro β é um factor exponencial que dá a probabilidade de distribuição das distâncias interatômicas, podendo ser interpretado como um factor de temperatura que define o movimento dos átomos. O parâmetro f é um factor de escala e A o número de átomos.

I-2.2.3: Descritores SURFACE

Os descritores SURFACE utilizam como *input* as propriedades físico-químicas calculadas pelo programa PETRA e um modelo 3D para a molécula e codificam a superfície de uma determinada molécula.⁸⁵ Nessa codificação são considerados todos os átomos de uma molécula e o volume de Van der Waals que cada átomo ocupa, sendo determinada a superfície acessível a uma esfera de solvente de raio r (**Figura I.17**).

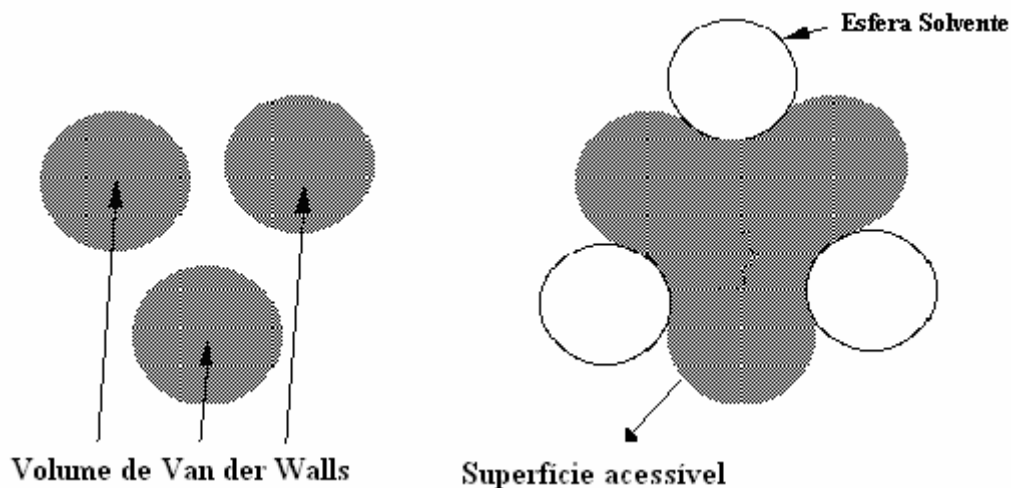


Figura I.17: Obtenção da superfície molecular tendo em conta o volume de Van der Waals ocupado por cada átomo da molécula e uma esfera de solvente de raio r .

Na prática esta superfície não é contínua mas é constituída por uma determinada densidade de pontos (**Figura I.18**).

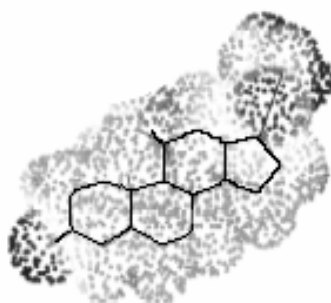


Figura I.18: Superfície molecular definida por uma determinada densidade de pontos.

O descritor SURFACE consiste num coeficiente de autocorrelação obtido tendo em conta o produto de propriedades físico-químicas (p_i, p_j) associadas a pares de pontos, sendo feito o somatório de todos os pares de pontos dentro de uma mesma gama de distâncias [$d_{inferior}, d_{superior}$] para o cálculo do coeficiente de autocorrelação respectivo:

$$A(d_{inferior}, d_{superior}) = \frac{1}{L} \sum_{ij} p_i p_j$$

Equação I.95

em que L é o número total de distâncias. Da mesma forma vão ser determinados outros coeficientes de correlação para outras gamas de distâncias, podendo as moléculas, independentemente do seu tamanho, estar representadas pelos mesmos tipos e número de coeficientes de autocorrelação.

I-2.2.4: Descritores XPET

É um conjunto de descritores moleculares cuja informação é extraída do resultado do programa PETRA, em que se incluem descritores globais como o peso molecular, número total de átomos ou o número de átomos de um determinado tipo. São incluídas também medidas de distribuição de cadeias, distribuição de cargas, carga máxima e carga mínima.

I-2.3: Descritores Chemaxon

A Chemaxon⁸⁶ é uma empresa de desenvolvimento e comercialização de *software* químico, entre os quais os *packages* JCHEMA e MARVINBEANS, ambos utilizados ao longo do trabalho apresentado nesta tese. O *package* MARVINBEANS permite a edição e visualização de estruturas químicas tendo incorporado um *plugin* (*cxcalc*) para cálculo de propriedades físico-químicas associadas a uma determinada molécula tanto no plano atómico como no plano global. Neste trabalho foram calculadas exclusivamente propriedades atómicas, sendo incorporadas na obtenção de descritores MOLMAP (detalhados no **capítulo III**). As propriedades atómicas determinadas para este trabalho são essencialmente baseadas em cargas σ , π e totais, electronegatividades associadas a ligações σ e π de um determinado átomo e polarizabilidades atómicas. Os métodos de cálculo são idênticos aos usados pelo programa PETRA muito embora, pela documentação disponibilizada pela Chemaxon, não se possa tirar conclusões relativamente ao método utilizado para determinação de cargas π e polarizabilidades. Por sua vez, o *package* JCHEMA foi criado pela Chemaxon para manipulação de bases de dados químicas, conversão entre diversos formatos de informação (.mol, .smi, .sdf), uniformização quanto à representação estrutural, obtenção de semelhanças entre diversas moléculas de uma base de dados e obtenção de descritores moleculares. No caso concreto deste trabalho este *package* foi utilizado para uniformizar representações estruturais, usando para tal o *plugin standardizer*.

I-3: Métodos de Selecção de Descritores

Os diversos métodos de aprendizagem automática já descritos têm diferentes sensibilidades quanto à presença de descritores não relevantes para a modelação/classificação da propriedade em estudo. Se as *Random Forest* operam bem com um número de descritores elevados apesar de alguns poderem ser eventualmente irrelevantes para o problema em questão, por outro lado as redes neuronais de *Kohonen* e CPG são sensíveis à presença de descritores irrelevantes – esses descritores podem

mesmo piorar muito significativamente a qualidade de um modelo. Para a selecção de descritores, de forma a eliminar descritores irrelevantes, foram utilizados neste trabalho algoritmos genéticos, que em seguida descrevemos.

Algoritmos Genéticos⁸⁷

O programa utilizado neste trabalho e que implementa os algoritmos genéticos é o SELVA. Este programa começa por ler dois ficheiros de *input*, um com o conjunto de *treino* e o outro com o conjunto de *previsão*. Cada objecto nos conjuntos de *treino* e *previsão* está representado por um conjunto de descritores moleculares. O objectivo do programa Selva é encontrar um conjunto mais reduzido de descritores, conjunto esse que permita a obtenção de um modelo o mais adequado possível à actividade/propriedade em estudo. Os algoritmos genéticos simulam a evolução de uma população, em que, cada indivíduo da população representa um sub-conjunto de descritores, cuja aptidão é avaliada através da capacidade de obtenção de modelos adequados. Assim cada indivíduo aparece representado pelo seu *cromossoma*, que tem tantos *genes* como o número de descritores iniciais. Determinado *gene* adquire o valor “1” se o respectivo descritor está incluído no sub-conjunto caso contrário adquire o valor “0” (Figura I.19).

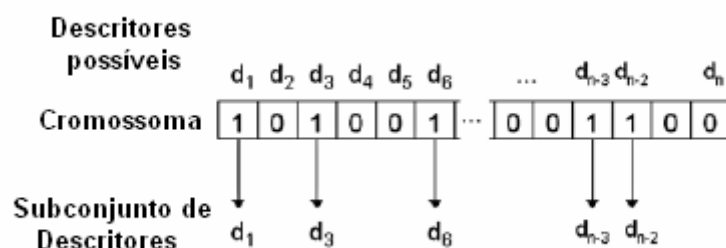


Figura I.19: Determinado cromossoma representa um sub-conjunto de descritores. O “0” significa ausência de descritor e “1” significa presença de descritor.

No início do processo evolutivo, os *cromossomas* são gerados aleatoriamente. Para que o número de descritores seleccionados seja mantido baixo (pequenos sub-conjuntos de descritores) a probabilidade de gerar “1’s” é mantida baixa (entre 0 e 0,4). A população de indivíduos é autorizada a evoluir durante determinado número de gerações. Em cada geração, metade da população morre e a outra metade sobrevive (os indivíduos mais aptos). Cada indivíduo que sobreviveu vai cruzar-se com outro indivíduo sobrevivente, resultando em cada cruzamento dois indivíduos descendentes. O cromossoma de cada indivíduo descendente resulta do cruzamento entre cromossomas parentais, *crossover*, seguido de mutação aleatória (Figura I.20):

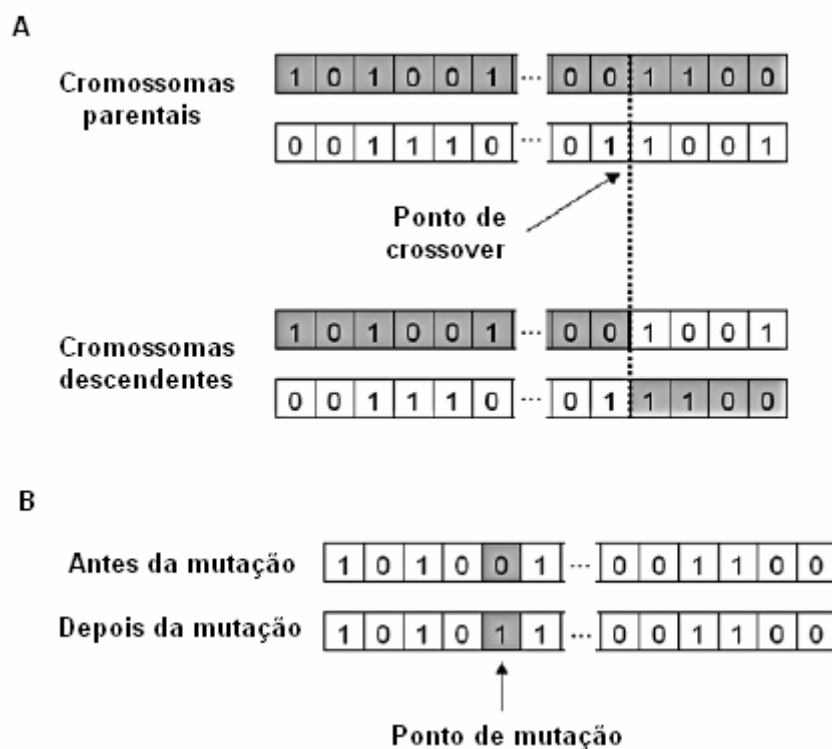


Figura I.20: A) Crossover entre cromossomas parentais, para obtenção dos cromossomas descendentes. B) Mutação de um gene.

A população da geração seguinte está constituída pelos indivíduos descendentes e os indivíduos parentais.

O *crossover* ocorre em determinado ponto aleatório. A mutação é autorizada a ocorrer em cada gene com uma probabilidade de ocorrência aleatória. A probabilidade de mutação $0 \rightarrow 1$ é determinada aleatoriamente para cada indivíduo entre 0 e 5%. A probabilidade de mutação $1 \rightarrow 0$ é 4 a 6 vezes maior que a probabilidade de mutação $0 \rightarrow 1$.

A avaliação de cada indivíduo é obtida através de uma rede neuronal de *Counter-Propagation* treinada com aquele subconjunto de descritores. A rede neuronal assim treinada constitui um modelo que vai ser usado na previsão da propriedade/atividade de interesse relativamente ao conjunto de previsão, sendo obtido o desvio entre valores previstos e valores experimentais, resultando o valor de RMS, que constitui uma medida de comparação entre indivíduos.

I-4: Bibliografia

- (1) <http://www.eisnt.com/>
- (2) Zupan, J.; Gasteiger J., *Neural Networks in Chemistry and Drug Design*, 2nd Ed., Wiley-VCH, Weinheim, **1999**.
- (3) Arakawa, M.; Hasegawa, K.; Funatsu, K.; *Chemom. Intell. Lab. Syst.*, **2006**, 83, 91-98.
- (4) Milinsek, G.; Novic M.; Hodoscek, M.; Solmajer, T.; *J. Chem. Inf. Comput. Sci.*, **2001**, 41, 1286-1294.
- (5) Satoh, H.; Funatsu, K.; Takano, K.; Nakata, T.; *Bull. Chem. Soc. Jpn.*, **2000**, 73, 1955-1965.
- (6) Gupta, S.; Matthew, S.; Abreu, P. M.; Aires-de-Sousa, J.; *Bioorg. Med. Chem.*, **2006**, 14, 1199-1206.
- (7) Carrera, G. V. S. M.; Branco, L. C.; Aires-de Sousa, J.; Afonso, C. A. M.; *Tetrahedron*, **2008**, 64, 2216-2224.
- (8) Carrera, G.; Aires-de-Sousa, J. *Green Chem.* **2005**, 7, 20-27.
- (9) Zupan, J.; Novic, M.; Gasteiger, J.; *Chemom. Intell. Lab. Sys.*, **1995**, 27, 175-187.
- (10) Peterson, K. L.; *Anal. Chem.*, **1992**, 64, 379-386.
- (11) Spycher, S.; Pellegrini, E.; Gasteiger, J.; *J. Chem. Inf. Model.*, **2005**, 45, 200-208.
- (12) Wu, C.; Chen, H-L.; Chen, S-C.; *Appl. Intell.*, **1997**, 7, 27-38.
- (13) Valkova, I.; Vracko, M.; Basak, S. C.; *Anal. Chim. Acta.*, **2004**, 509, 179-186.
- (14) Panek, J. J.; Jezierska, A.; Vracko, M.; *J. Chem. Inf. Model.*, **2005**, 45, 264-272.
- (15) Gramatica, P.; Navas N.; Todeschini, R.; *Trends Anal. Chem.*, **1999**, 18, 461-471.
- (16) Peterson, K. L.; *J. Chem. Inf. Comput. Sci.*, **1995**, 35, 896-904.
- (17) Chen, L.; Gasteiger, J.; *J. Am. Chem. Soc.*, **1997**, 119, 4033-4042.
- (18) Zhang, Q.; Aires-de-Sousa, J.; *J. Chem. Inf. Model.*, **2005**, 45, 1775-1783.
- (19) Latino, D. A. R. S.; Aires-de-Sousa, J.; *Angew. Chem. Int. Ed.*, **2006**, 45, 2066-2069.
- (20) Simon, V.; Gasteiger, J.; Zupan, J.; *J. Am. Chem. Soc.*, **1993**, 115, 9148-9159.
- (21) da Fonseca, A. M. P.; Dissertação de Mestrado “Classificação Geográfica Automática de Petróleos Brutos”, UNL, Lisboa, **2005**
- (22) Wang, Y. H.; Li, Y.; Yang, S. L.; Yang, L.; *J. Chem. Inf Model.*, **2005**, 45, 750-757.
- (23) Gasteiger, J.; Teckentrup, A.; Terfloth, L.; Spycher, S.; *J. Phys. Org. Chem.*, **2003**, 16, 232-245.
- (24) Niculescu, S. P.; *Tetrahedron*, **2003**, 622, 71-83.
- (25) Briollais, L.; Wang, Y.; Rajendram, I.; Onay, V.; Shi, E.; Knight, J.; Ozcelik, H.; *BMC Med*, **2007**, 5, 22-37.
- (26) Hill, T.; Lewicki, P.; *Statistics: Methods and Applications : a Comprehensive Reference for Science, Industry, and Data Mining*, Ed. StatSoft, **2006**.
- (27) Breiman, L.; Friedman, J. H.; Olshen, R. A.; Stone, C. J.; *Classification and Regression Trees*, Chapman & Hall/CRC, Boca Raton, Florida, **2000**.
- (28) Daszykowski, M.; Walczak, B.; Xu, Q. -S.; Daeyaert, F.; de Jonge, M. R.; Heeres, J.; Koymans, L. M. H.; Lewi, P. J.; Vinkers, H. M.; Janssen, P. A.; Massart, D. L.; *J. Chem. Inf. Comput. Sci.*, **2004**, 44, 716-726.
- (29) Caetano, S.; Aires-de-Sousa, J.; Daszykowski, M.; Heyden, Y. V., *Anal. Chim. Acta*, **2005**, 544, 315-326.
- (30) Arena, V. C.; Sussman, N. B.; Mazumdar, S.; Yu, S.; Macina, O. T.; *SAR QSAR Environ. Res.*, **2004**, 15, 1-18.
- (31) Markey, M. K.; Tourassi, G. D.; Floyd Jr., C. E.; *Proteomics*, **2003**, 3, 1678-1679.
- (32) Lee, A. C.; Yu, J. Y.; Crippen, G. M.; *J. Chem. Inf. Model*, **2008**, 48, 2042-2053.
- (33) Schneider, N.; Jackels, C.; Andres, C.; Hutter, M. C.; *J. Chem. Inf. Model*, **2008**, 48, 613-628.
- (34) Grindon, C.; Combes, R.; Cronin, M. T. D.; Roberts, D. W.; Garrod, J. F., *Alt. Lab. Anim.*, **2008**, 36, 65-85.
- (35) Huang, Y. Y.; Tseng, G. C.; Yuan, S. S.; Pasa-Tolic, L.; Lipton, M. S.; Smith, R. D.; Wysochi, V. H.; *J. Prot. Res.*, **2008**, 7, 70-79.

- (36) Lin, H. H.; Han, L. Y.; Yap, C. W.; Xue, Y.; Liu, X. H.; Zhu, F.; Chen, Y. Z.; *J. Mol. Graph. Model.*, **2007**, 26, 505-518.
- (37) Neugbauer, A.; Hartmann, R. W.; Klein, C. D.; *J. Med. Chem.*, **2007**, 50, 4665-4668.
- (38) Andres, C.; Hutter, M. C.; *QSAR Comb. Sci.*, **2006**, 25, 305-309.
- (39) Kim, H. J.; Choo, H.; Cho, Y. S.; Koh, H. Y.; No, K. T.; Pae, A. N.; *Bioorg. Med. Chem.*, **2006**, 14, 2763-2770.
- (40) Manga, N.; Duffy, J. C.; Rowe, P. H.; Cronin, M. T. D.; *SAR QSAR Environ. Res.*, **2005**, 16, 43-61.
- (41) Breiman, L.; *Machine Learning*, **2001**, 45, 5-32.
- (42) Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J. C.; Sheridan, R. P.; Feuston, B. P.; *J. Chem. Inf. Comput. Sci.*, **2003**, 43, 1947-1958.
- (43) Li, S.; Fedorowicz, A.; Singh, H.; Soderholm, S. C.; *J. Chem. Inf. Model*, **2005**, 45, 952-964.
- (44) Zhang, Q.; Aires-de-Sousa, J.; *J. Chem. Inf. Model.*; **2007**, 47, 1-8.
- (45) Palmer, D. S.; O'Boyle, N. M.; Glen, R. C.; Mitchell, J. B. O.; *J. Chem. Inf. Model.*, **2007**, 47, 150-158.
- (46) Cannon, E. O.; Bender, A.; Palmer, D. S.; Mitchell, J. B. O.; *J. Chem. Inf. Model.*, **2006**, 46, 2369-2380.
- (47) Hughes, L. D.; Palmer, D. S.; Nigsch, F.; Mitchell, J. B. O.; *J. Chem. Inf. Model.*, **2008**, 48, 220-232.
- (48) Blower, P. E.; Cross, K. P.; Eichler, G. S.; Myatt, G. J.; Weinstein, J. N.; Yang, C.; *Comb. Chem. High Throughput Screen.*, **2006**, 9, 115-122.
- (49) Sheridan, R. P.; Korzekwa, K. R.; Torres, R. A.; Walker, M. J.; *J. Med. Chem.*, **2007**, 50, 3173-3184.
- (50) Khandelval, A.; Bahadduri, P. M.; Chang C.; Polli, J. E.; Swaan, P. W.; Ekins, S.; *Pharm. Res.*, **2007**, 24, 2249-2262.
- (51) Wang, H.; Klinginsmith, J.; Dong, X.; Lee, A. C.; Guha, R.; Wu, Y.; Crippen, G. M.; Wild, D. J.; *J. Chem. Inf. Model.*, **2007**, 47, 2063-2076.
- (52) Nandigam, R. K.; Evans, D. A.; Erickson, J. A.; Kim, S.; Sutherland, J. J.; *J. Chem. Inf. Model.*, **2008**, 48, 2386-2394.
- (53) Debeljak, Z.; Skrbo, A.; Jasprica, I.; Mornar, A.; Plecko, V.; Banjanac, M.; Medic-Saric, M.; *J. Chem. Inf. Model.*, **2007**, 47, 918-926.
- (54) Song, M.; Clark, M.; *J. Chem. Inf. Model.*, **2006**, 46, 392-400.
- (55) Steinwart, I.; *J. Complexity*, **2002**, 18, 768-791.
- (56) Kriegl, J. M.; Arnhold, T.; Beck, B.; Fox, T.; *QSAR Comb. Sci.*, **2005**, 24, 491-502.
- (57) Cai, Y-D.; Liu, X-J.; Xu, X-B.; Chou, K-C.; *Comput. Chem.*, **2002**, 26, 293-296.
- (58) Cai, Y-D.; Liu, X-J.; Xu, X-B.; Chou, K-C.; *Comput. Chem.*, **2002**, 23, 267-274.
- (59) Luan, F.; Zhang, R.; Zhao, C.; Yao, X.; Liu, M.; Hu, Z.; Fan, B.; *Chem. Res. Toxicol.*, **2005**, 18, 198-203.
- (60) Caetano, S.; Ustun, B.; Hennessy, S.; Smeyers-Verbeke, J.; Meissen, W.; Downey, G.; Buydens, L.; Heyden, Y. V.; *J. Chemom.*, **2007**, 21, 324-334.
- (61) Yan, G. K.; Li, J. J.; Li, B. R.; Hu, J.; Guo, W. P.; *J. Theory Comput. Chem.*, **2007**, 6, 495-509.
- (62) Qin, S.; Liu, H. X.; Wang, J.; Yao, X. J.; Liu, M. C.; Hu, Z. D.; Fan, B. T.; *QSAR Comb. Sci.*, **2007**, 26, 443-451.
- (63) Liao, Q.; Yao, J. H.; Yuan, S. G.; *Mol. Divers.*, **2006**, 10, 301-309.
- (64) Chen, N. Y.; Ding, Y. M.; Bao, X. H.; Liu, X.; Lu, W. C.; *J. Rare Earth.*, **2005**, 23, 30-34.
- (65) Frohlich, H.; Wegner, J. K.; Zell, A.; *QSAR Comb. Sci.*, **2004**, 23, 311-318.
- (66) Lind, P.; Maltseva, T.; *J. Chem. Inf. Comput. Sci.*, **2003**, 43, 1855-1859.
- (67) Dudek, A. Z.; Arodz, T.; Gálvez, J.; *Comb. Chem. High Throughput Screen.*, **2006**, 9, 213-228.
- (68) Xu, Y.; Zomer, S.; Brereton, R. G.; *Crit. Rev. Anal. Chem.*, **2006**, 36, 177-188.
- (69) Gunn, S. R.; *Support Vector Machines for Classification and Regression*, University of Southampton, **1998**.
- (70) Smola, A. J.; Scholkopf, B.; *Stat. Comput.*, **2004**, 14, 199-222.
- (71) Mais informações sobre empresa TALETE em <http://www.taletе.mi.it/index.htm>.

- (72) Cálculo de descritores *online* em: <http://www.vcclab.org/lab/edragon/start.html>.
- (73) consultar página web: <http://www.chemistry.ohio-state.edu/~woodward/ch754/electro.htm>
- (74) Consultar definições na página web: <http://www.ra.cs.uni-tuebingen.de/software/joelib/tutorial/descriptors/descriptors.html>
- (75) Randic, M.; *J. Chem. Inf. Comput. Sci.*, **2001**, 41, 607-613.
- (76) Araújo, O.; De la Peña, J. A.; *J. Chem. Inf. Comput. Sci.*, **1998**, 38, 827-831.
- (77) Petitjean, M.; *J. Chem. Inf. Comput. Sci.*, **1992**, 32, 331-337.
- (78) Sharma, V.; Goswami, R.; Madan, A. K.; *J. Chem. Inf. Comput. Sci.*, **1997**, 37, 273-282.
- (79) Diversas definições de descritores moleculares em:
http://www.strandls.com/sarchitect/documents/manual_html/desctheory.html
- (80) Definições destes e de outros descritores em: <http://www.codessa-pro.com/descriptors/topo/bici.htm>.
- (81) Manual do PETRA em <http://www2.ccc.uni-erlangen.de/software/petra/manual/>.
- (82) Kleinoder, T; “*Prediction of Properties of Organic Compounds. Empirical Methods and Management of Property Data*”, Friedrich Alexander Universität, Erlangen-Nürnberg, **2006**.
- (83) Gasteiger, J.; Saller, H.; *Angew Chem. Int. Ed. Engl.*, **1985**, 24, 687-689.
- (84) Gasteiger, J.; A Hierarchy of Structure Representations; vol. 3 in *Handbook of Chemoinformatics*, eds Gasteiger, J., Engel, T., Wiley-VCH, Weinheim, **2003**.
- (85) Wagner, M.; Sadowski, J.; Gasteiger, J.; *J. Am. Chem. Soc.*, **1995**, 117, 7769.
- (86) Consultar página web: <http://www.chemaxon.com/>.
- (87) Homeyer, A. V. Evolutionary Algorithms and their Applications in Chemistry. *Handbook of Chemoinformatics*, eds Gasteiger, J., Engel, J.; Wiley-VCH: Weinheim, **2003**.

Capítulo II

Estudo de propriedades físico-químicas de líquidos iônicos

II-1: Introdução Teórica

II - 1.1: Enquadramento

No contexto da sociedade globalizada em que vivemos onde a competitividade é cada vez mais um factor fundamental e onde a sustentabilidade do nosso planeta está na ordem do dia pelos piores motivos possíveis, surge a necessidade de novos processos de produção mais eficientes e ao mesmo tempo mais sustentáveis ao nível ambiental. Nesse contexto os líquidos iónicos apresentam-se como uma alternativa potencialmente favorável ao desenvolvimento de processos mais eficazes e ao mesmo tempo mais sustentáveis. Os líquidos iónicos são uma nova classe de compostos, que, tal como o próprio nome indica, são constituídos por iões (normalmente um catião e um anião) e são líquidos à temperatura ambiente ou a uma temperatura próxima (ponto de fusão inferior a 100°C). Esta nova classe de compostos possui como uma das suas características mais atraentes a sua quase negligenciável volatilidade, conduzindo a quase nulas emissões para atmosfera, facto que os torna em candidatos a substituir os tradicionais solventes orgânicos voláteis usados em síntese orgânica e não só. Para além da baixa volatilidade, os líquidos iónicos podem ver as suas propriedades físico-químicas modeladas e adaptadas a uma determinada situação específica pela escolha judiciosa do catião e do anião, sendo considerados materiais avançados com um âmbito de aplicação muito para além do seu uso como solventes em síntese. As 10^8 possibilidades de combinação entre catião e anião¹ ao mesmo tempo que oferecem a possibilidade de obtenção de um material com as características pretendidas como também constituem um grande desafio na hora de escolher a combinação certa para uma dada situação. A procura do material com as características adequadas simplesmente pelo método da tentativa e erro é impraticável tanto ao nível de tempo como económico, logo estudos detalhados sobre a relação entre a estrutura e propriedades físico-químicas são condição *sine qua non* para um conhecimento profundo sobre esta classe de compostos e para poder dirigir a sua síntese e assim poupar tempo e recursos.

Propriedades como o ponto de fusão e densidade são propriedades fundamentais para definir o âmbito de aplicação de um dado líquido iónico, e a relação com a sua estrutura é complexa. Nesse contexto o recurso a métodos de aprendizagem automáticos não lineares, como é o caso de Redes Neurais,² *Random Forest*,³ *Support Vector Machines*^{4,5} ou Árvores de Decisão,⁶ que têm sido usados amplamente na obtenção de modelos para a previsão de inúmeras propriedades físico-químicas e biológicas, adequa-se à modelação da densidade e ponto de fusão de líquidos iónicos. Os métodos de aprendizagem automática utilizados são complementares, tanto na obtenção de informação relevante sobre a propriedade a modelar e relação com a sua estrutura, como também na selecção de variáveis estruturais relevantes para obtenção dos mencionados modelos.

Neste trabalho, os modelos obtidos foram utilizados para prever propriedades físico-químicas de novos sais e assim poder aceder *à priori* ao valor dessas propriedades para determinada estrutura. Com essa informação foram sintetizados novos sais com as propriedades pretendidas. Estes métodos, pela rapidez e informação obtida, têm condições para poder vir a ser implementados por químicos orgânicos e não só, no sentido de concentrar esforços na síntese de novos líquidos iônicos com as propriedades físico-químicas pretendidas.

II-1.2: Líquidos Iônicos

II - 1.2.1: Aspectos Gerais de Líquidos Iônicos

Os líquidos iônicos, objecto de estudo neste capítulo da tese, apresentam como característica fundamental serem sais. Normalmente os sais estão associados a sólidos de elevado ponto de fusão como é o caso do cloreto de sódio; os líquidos iônicos (LIs) vêm derrubar esse mito, visto muitos serem líquidos à temperatura ambiente ou possuem um ponto de fusão inferior a 100°C, tecto de temperatura habitualmente aceite na designação de líquido iónico. Estes sais são constituídos por um catião, habitualmente orgânico, e por um anião orgânico ou inorgânico.

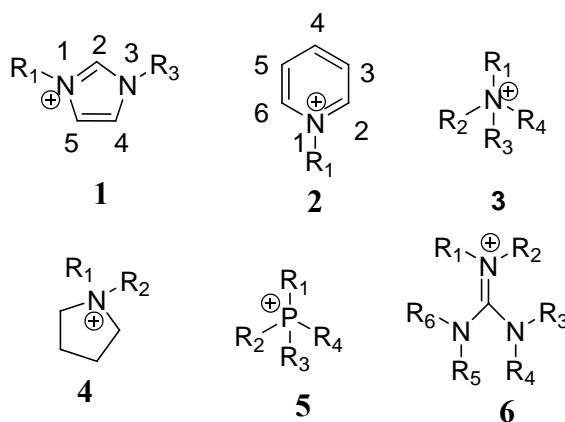


Figura II.1: Estruturas catiónicas habituais em líquidos iônicos.

Entre as famílias de catiões mais populares (**Figura II.1**) encontram-se os imidazólios **1**, piridínios **2**, amónios **3**, pirrolidínios **4**, fosfónios **5**, guanidínios **6**, entre outros.

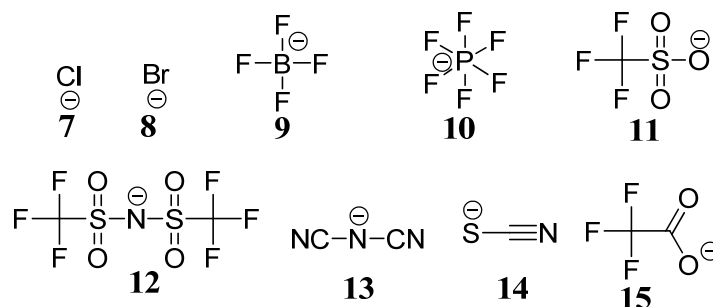


Figura II.2: Estruturas aniónicas habituais em líquidos iónicos.

Cloreto **7**, brometo **8**, tetrafluoroborato **9**, hexafluorofosfato **10**, trifluorometanosulfonato **11**, bis(trifluorometilsufonil)imidato **12**, dicianamida **13**, tiocianato **14** e trifluoroacetato **15**, são os aniões (**Figura II.2**) habitualmente mais usados em líquidos iónicos. Factores como a assimetria dos iões e deslocalização da carga são factores cruciais na obtenção de sais líquidos à temperatura ambiente (ou temperatura próxima da temperatura ambiente), por outras palavras, líquidos iónicos LIs.

Ainda que tenham aparecido recentemente na literatura científica como uma nova classe de compostos com aplicações nas mais variadas áreas, o conceito de LI não é novo, sendo necessário recuar até 1914 para ter descritos os primeiros LIs como é o caso do nitrato de etilamónio, descrito por Walden,⁷ líquido à temperatura ambiente (com quantidades de água entre 200 e 600 ppm).

Os primeiros líquidos iónicos baseados em aniões cloroaluminatos surgiram em 1948 pela mão de Hurley e Wier⁸ que usaram estes líquidos iónicos (piridínios) como meio electrólito para redução de catiões de alumínio para uso em revestimento.

Durante a década de 70 e início da década de 80, a investigação deste tipo de sistemas ganhou novo ânimo quando Osteryoung⁹ e Wilkes¹⁰ os redescobriram preparando com sucesso novos sais de cloroaluminatos de piridínio e imidazólio respectivamente, líquidos à temperatura ambiente, usados em estudos electroquímicos.

Na década de 80, os líquidos iónicos começaram a ganhar maior relevo sendo estudada a estabilidade de certos metais de transição. Seddon¹¹ iniciou o seu importante contributo na área dos LIs, estudando concretamente a estabilidade de complexos de molibdénio na presença de sais cloroaluminatos de piridínio e imidazólio. Este autor foi um dos principais impulsionadores dos LIs.

Foi já na segunda metade da década de 80 que o grupo de Wilkes¹² usou pela primeira vez líquidos iónicos, mais concretamente cloroaluminatos de imidazólio, como meio reaccional e catalisador simultaneamente em alquilações e acilações de Friedel Crafts. A partir daí, os líquidos iónicos começaram a despertar o interesse em catálise sendo usados por Chauvin *et al*,¹³ já na década

de 90, na dimerização do propeno, usando para tal cloroaluminatos de imidazólio, piridínio e fosfónio.

Foi no início da década de 90 que o conceito de LI recebeu o impulso mais significativo para o interesse no seu estudo, quando o grupo de Wilkes,¹⁴ preparou LIs estáveis na presença de água como é o caso de sais de tetrafluoroborato. A partir dessa observação a gama de aplicações de líquidos iónicos em síntese orgânica e não só, aumentou muito significativamente, sendo actualmente uma área consideravelmente “quente” no que concerne à produção científica.

II - 1.2.1.1: Líquidos Iónicos em Síntese Orgânica

Os LIs têm sido usados com sucesso em inúmeras aplicações, nomeadamente em síntese orgânica¹⁵ em reacções de hidrogenação catalítica onde sais de imidazólio, na presença de catalisadores de Ródio e Ruténio, permitiram a obtenção de velocidades de reacção mais elevadas que em solventes orgânicos tradicionais; observou-se elevados rendimentos, elevadas selectividades e enantioselectividades. O âmbito de aplicação das hidroformilações foi alargado a alcenos de cadeias maiores usando líquidos iónicos como meio reaccional nomeadamente sais de imidazólio e amónio. Dimerização, oligomerização e polimerização de olefinas com índices de produção mais elevados foram obtidos na presença de líquidos iónicos, estes resultados conduziram à primeira aplicação em escala industrial de LIs, concretamente no processo de dimerização de olefinas denominado Difasol licenciado para a IFP.¹⁶ Em reacções de polimerização de olefinas observou-se que o tamanho da cadeia alquílica em sais 1-alkil-3-metilimidazólio e *N*-alkilpiridínio influencia diferencialmente a actividade catalítica e o grau de polimerização, sendo este um claro exemplo de uso de líquidos iónicos à medida das necessidades.¹⁵ Continuando a percorrer a lista das reacções orgânicas em que o uso de líquidos iónicos se tornou numa mais valia, encontra-se a reacção de Heck verificando-se incrementos nos rendimentos obtidos em líquidos iónicos comparativamente a outros solventes orgânicos como é o caso de DMF. Sais de imidazólio e amónio figuram entre os casos de sucesso relativamente a esta reacção. Outros casos de sucesso foram relatados¹⁷ como o caso da reacção de Stille, adição de Michael, condensação de Aldol, Baylis-Hillman, reacção de Wittig e monoalkilação de aminas primárias.¹⁸ Relativamente a reacções de Diels-Alder, o uso de sais de imidazólio como meio reaccional permitiu uso de menores quantidades de catalisador e melhorar a selectividade endo/exo. Ainda relativamente às reacções de Diels-Alder, mas usando LIs (sais de imidazólio) cristais líquidos foi possível ter uma selectividade exo/endo 54/46 quando em etanol essa selectividade é de 12/88.¹⁹

II - 1.2.1.2: Líquidos Iónicos como Cristais Líquidos e Outros Materiais

O uso dos líquidos iónicos vai muito para além da sua aplicação em síntese orgânica. Podem ser também materiais avançados para aplicações específicas, como é o caso de cristais líquidos, meios em que há mobilidade das moléculas mas ao mesmo tempo as moléculas apresentam ordem, essa ordem pode ser ao nível da posição ou só ao nível da orientação. Essa ordem pode-se alterar com a temperatura, ou seja, é possível ter cristais líquidos termotrópicos. Assim, no caso genérico de um sólido cristalino em que se vai aumentando a temperatura, irá aparecer um ponto de fusão, formando-se uma fase líquida cristalina cuja ordem é posicional (fase esmética), continuando a aumentar a temperatura vai haver outra mudança de fase, neste caso para uma fase cristalina cuja ordem se prende apenas com a orientação das moléculas (fase nemática), continuando a aumentar a temperatura as moléculas acabarão por perder ordem constituindo a transição para a fase isotrópica; para um determinado cristal líquido pode acontecer só se observar uma ou diversas fases líquidas cristalinas (mesofases). Cristais líquidos são habitualmente usados em LCDs muito embora a existência de cristais líquidos que ao mesmo tempo são líquidos iónicos aumentou o âmbito de aplicação destes materiais. Estes compostos foram usados como materiais de condutividade centrada em iões de carácter anisotrópico, bem como meio reaccional ordenado para realizar reacções químicas com selectividades distintas das observadas em solventes mais tradicionais.^{19,20} Estruturalmente foram encontrados cristais líquidos em sais de amónio (alguns deles sendo também surfactantes) a sua grande maioria com pelo menos uma cadeia alquílica longa (normalmente acima de 10 átomos de carbono). Esses sais de amónio geralmente têm como aniões, brometos, cloretos, ácidos aromáticos (na forma desprotonada), aniões metaloclorados ($[MCl_4]^{-2}$) de zinco, cádmio, mercúrio, cobre, paládio, cobalto, ferro, níquel e chumbo entre outros correspondendo a cristais líquidos com estabilidades e gamas de temperatura em que existe a fase líquida cristalina variáveis. A natureza das fases cristalinas pode também ser modelada pela presença de determinadas funcionalidades no catião, como é o caso dos grupos nitrilo. Relativamente a fosfónios estes apresentam características muito semelhantes aos dos sais de amónio principalmente no que toca à relação entre estrutura e natureza da(s) mesofase(s), havendo algumas subtis diferenças entre sais de amónio e fosfónio. Está descrita a preparação de cloretos de di-n-alkil fosfónio (cadeias alquílicas C_{10} , C_{14} e C_{18}) apresentando estes gamas de temperatura, em que existe a mesofase, e estabilidades superiores aos correspondentes sais de amónio.²⁰ Relativamente a sais de imidazólio, o seu estudo térmico está maioritariamente associado a sais de 1-alkil-3-metilimidazólio, tendo sido encontrado comportamento de cristal líquido quando o grupo alkilo normalmente é superior a C_{12} . Com este tipo de catiões, foi encontrado comportamento líquido cristalino na presença de aniões cloreto, metalohalogenados ($[MCl_4]^{-2}$) de cobalto, níquel e paládio, hexafluorofosfato²¹ e tetrafluoroborato.²² Observando-se nestes dois últimos aniões um aumento da gama de temperaturas em que existe uma fase líquida cristalina com o aumento da cadeia alquílica do 1-alkil-3-metilimidazólio. Ficou demonstrado que alguns destes compostos

têm uma condutividade iônica muito elevada.²⁰ Também se demonstrou que não é necessário ter uma cadeia muito longa nos 1-alkil-3-metilimidazólios contanto que a mesma exista no anião, como é o caso do dodecilsulfato de 1,3-dimetilimidazólio que apresenta comportamento de cristal líquido. Relativamente a este caso ainda se observou que a presença ou não de um grupo metilo adicional no anel de imidazólio pode controlar a formação ou não de cristais líquidos, ou até mesmo a total perda de cristalinidade,²⁰ ficando mais uma vez evidente a capacidade de um líquido iônico ser concebido à medida de um determinado propósito modificando parte da sua estrutura. Ainda que muito menos estudados, os sais de 1,3-dialquilimidazólio, em algumas situações, apresentaram fase líquida cristalina, nomeadamente na presença de aniões de cloreto, brometo, hexafluorofosfato e aniões clorometalados ($[MCl_4]^{-2}$). Recentemente foram caracterizados novos sais de imidazólio baseados no anião *para*-toluenosulfonato exibindo comportamento de cristal líquido.²³

Os sais de piridínio e a sua aplicação como cristais líquidos começou a ser estudada no longínquo ano de 1938.²⁰ À semelhança de alguns sais de imidazólio, verificou-se que para cloretos, brometos e iodetos de *N*-(*n*-alkil)-piridínio com cadeia alquílica a partir de C_{12} o aumento da cadeia provoca um aumento da estabilidade da mesofase. Posteriormente aniões metaloclorados foram usados em combinação com este tipo de catiões na obtenção de mesofase. Presença de longos grupos alquilo noutras posições do anel piridínio também permitiram a obtenção de mesofases, nomeadamente nas posição 2- ou 4-. Neste último caso a presença de grupo metilo na posição 1- vai permitir a formação de mesofase, enquanto que protonação ou presença de outros grupos alquilo nessa posição conduzem à inexistência de mesofase. Halogenetos foram usados nestes estudos. Estudou-se também o efeito da ramificação em grupos alquilo na existência e natureza da mesofase verificando-se uma maior desestabilização quando o ponto de ramificação está próximo do anel de piridínio, no mesmo sentido a substituição de grupo metilo na posição 4- provoca aumento de ponto de fusão e desestabilização da mesofase. Já a presença de posição olefínica terminal na cadeia alquílica de alguns sais de piridínio conduziu a uma temperatura de passagem para fase isotrópica mais baixa. Em alguns casos também se verificou que o aumento do tamanho do halogeneto como anião vai provocar decréscimo no espaçamento entre as camadas na mesofase, provocando um decréscimo na gama de temperaturas em que aparece mesofase. Os sais de guanidínio, por sua vez, foram pouco estudados no âmbito dos cristais líquidos, ainda assim dois estudos foram efectuados. Um estudo foi feito por Mathevet *et al*²⁴ com alcanosulfonatos de guanidínio **16** (**Figura II.3**), em que fase líquida cristalina esméctica foi observada a partir de cadeias C_8 até cadeias C_{18} no anião, havendo transição para uma fase esméctica ordenada a temperaturas de 63°C para C_8 e 97°C para C_{18} .

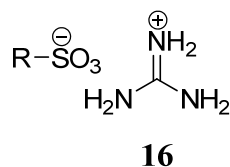


Figura II.3: Estrutura genérica de um alcanosulfonato de guanidínio.

Outro estudo, realizado por Martin et al.²⁵ utilizando alquilbenzosulfonatos **17** e alquilbifenilsulfonatos **18** de guanidínio (**Figura II.4**) em que foi detectada fase líquida cristalina a partir de cadeias C₇ até C₁₆ nos aniões. Transições para fase esmética foram observadas entre 120-180°C para alquilbenzosulfonatos de guanidínio **17**. Relativamente aos alquilbifenilsulfonatos de guanidínio **18** a transição para fase esmética deu-se entre 80-220°C havendo em geral, e à semelhança dos alquilbenzosulfonatos **17**, decréscimo na temperatura de transição com o aumento da cadeia alquílica. Esta família de compostos apresenta em geral transições a temperaturas mais elevadas que os correspondentes alquilbenzosulfonatos **17**.

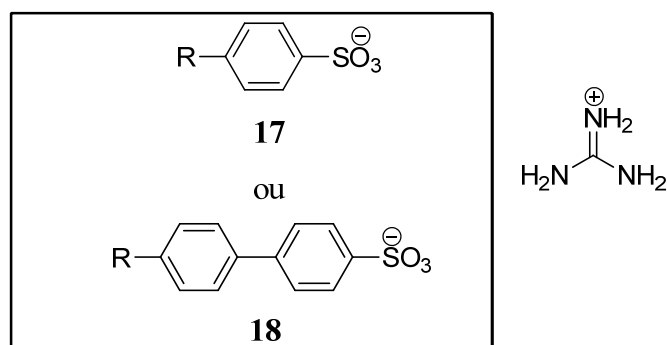


Figura II.4: Estrutura genérica de alquilbenzosulfonatos **17 e alquilbifenilsulfonatos **18** de guanidínio.**

O uso de líquidos iônicos como materiais avançados não se limita ao seu uso como cristais líquidos. Os líquidos iônicos permitem largo espectro de potencial electroquímico, tal como elevadas condutividades eléctricas, características que associadas à sua capacidade de dissolver compostos orgânicos, inorgânicos e organometálicos e largo espectro de existência no estado líquido fazem dos líquidos iônicos excelentes meios electrolitos.²⁶ Líquidos iônicos são também usados em membranas líquidas, separando eficazmente gases e moléculas orgânicas (também com interesse biológico), podendo actuar como sensores, nomeadamente de oxigénio, ou como reactores de membrana, facilitando a separação entre produtos e reagentes. Características como a baixa volatilidade e capacidade de ajustar as suas capacidades de permeabilidade fazem dos líquidos iônicos os substitutos ideais das tradicionais membranas líquidas.²⁶ Na área analítica, os líquidos iônicos têm sido usados com sucesso como fase estacionária (quiral e não quiral) em cromatografia gasosa e usados como aditivos na fase móvel em HPLC facilitando separação de compostos. Têm sido usados com sucesso

como meio electrólito em electroforese capilar, sendo também usados como matrizes no método MALDI de espectrometria de massa.²⁶ Líquidos iónicos podem ser usados como agentes dispersantes em tintas e pigmentos, estabilizando-os, reduzindo a sua viscosidade e eliminando a sedimentação. Podem também actuar como surfactantes, baixando a tensão interfacial, permitindo impregnação da superfície, solubilização, emulsificação e dispersão.²⁶

Alguns líquidos iónicos apresentam actividade anti-microbiana, principalmente aqueles possuindo grupos alquilo longos ou mesmo grupos alcóxido com cadeia alquílica longa, podendo ser usados como agentes de embalsamamento, ou como aditivos a usar em papel para evitar degradação de documentos.²⁶ Em adição, os líquidos iónicos podem ser usados como agentes de revestimento anti-corrosivo, nomeadamente no aço.²⁶

Na área da nanotecnologia o recurso a líquidos iónicos como meio reaccional organizado, com reduzida tensão superficial permite a síntese de nanopartículas e outras nanoestruturas com propriedades melhoradas e ajustadas a uma particular aplicação modeladas pela estrutura do líquido iónico. A título de exemplo nanopartículas de Iridio e Ruténio muito finas e estáveis foram sintetizadas por redução em líquidos iónicos.²⁷ O emprego de líquidos iónicos na síntese super-estruturas cristalinas de dióxido de titânio, permitiu uma simplificação dos passos efectuados na sua síntese, recorrendo a quantidades mínimas de água, sendo obtida uma melhoria efectiva na qualidade destas estruturas que podem ser usadas na conversão de energia solar em células solares.²⁷ Verificou-se também que as propriedades hidrofílicas e hidrofóbicas de nanopartículas de ouro podiam ser modeladas pela alteração do anião no líquido iónico usado na síntese.²⁷

Ainda na área dos materiais, Stracke *et al*²⁸ desenvolveram líquidos iónicos centrados na unidade imidazólio com substituinte o ciclo-hexano, constituindo um meio pouco volátil, não inflamável, denso, estável química e termicamente para o armazenamento de hidrogénio, neste caso como ligações C-H no ciclo-hexano, que na presença de catalisador de Paládio conduz à libertação de H₂ com a consequente formação de derivado benzénico. Processo inverso, ou seja, hidrogenação do derivado benzénico para dar origem a derivado ciclo-hexano, também é possível na presença de H₂ e catalisador de Paládio.

II - 1.2.1.3: Líquidos Iónicos em Extracções

Como método de separação energeticamente eficiente e selectivo, a extracção é amplamente usada por engenheiros químicos e químicos orgânicos. Na constante optimização deste processo os líquidos iónicos têm uma importante palavra a dizer, principalmente no que concerne ao nível da segurança, ambiente e selectividade. Líquidos iónicos foram usados com sucesso como meio hidrofóbico para a extracção de iões metálicos, por exemplo em meio aquoso. Para iões metálicos passarem para a fase do líquido iónico normalmente é necessário um agente quelante como os éteres

de coroa. A dissolução deste tipo de compostos é termodinamicamente favorecida em líquidos iónicos.²⁹ Desta forma, foi possível extrair com sucesso iões de metais alcalinos, alcalino-terrosos, metais pesados e radioactivos. Ainda assim foi adicionada complexidade ao sistema pelo recurso a quelantes que depois são difíceis de extrair do líquido iónico. Para contornar esse problema recorreu-se a TSILs (líquidos iónicos para uma aplicação específica), líquidos iónicos contendo funcionalidades específicas, que pudessem funcionar como agente quelante para um dado ião metálico. Foram usados com sucesso líquidos iónicos funcionalizados com grupos ureia e tioureia para a extracção de Hg^{2+} e Cd^{2+} . Relativamente a moléculas orgânicas, com interesse biológico ou com interesse como biocombustível, estas foram extraídas com sucesso para líquidos iónicos, principalmente provenientes de sistemas aquosos. Concretamente foram extraídos compostos fenólicos, aminoácidos, hidratos de carbono, ácidos orgânicos, antibióticos, hidrocarbonetos e biocombustíveis como é o caso do butanol.²⁹ Líquidos iónicos foram utilizados com sucesso na desulfurização de combustíveis, evitando o impacto na atmosfera dos SO_x .²⁹ Relativamente a gases, líquidos iónicos foram utilizados para extrair CO_2 com grau de solubilidade modelado pela presença de funcionalidade específica no catião de imidazólio como é o caso do grupo amina (TSIL), ou pela modificação do anião. Em todos estes processos de extracção o comportamento do líquido iónico pode ser modelado não só através da sua estrutura mas também pelo ajustamento de outras variáveis externas como é o caso do pH ou uso de sais inorgânicos para favorecer a separação de fases entre líquido iónico e meio aquoso.

Para isolar certos compostos do próprio líquido iónico usado para os extrair, pode-se recorrer à destilação ou à extracção com água. No caso de não ser possível usar nenhum destes métodos, pelo composto não ser volátil, por se degradar facilmente, reagir com a água, ou por não interessar de maneira nenhuma isolá-los em meio aquoso, a utilização do CO_2 super-crítico pode ser a solução, formando uma segunda fase na presença de líquidos iónicos polares. A sua volatilidade e possibilidade de modelar a solubilidade de compostos orgânicos simplesmente pela pressão empregue, fazem do CO_2 super-crítico um meio extractivo de eleição.²⁹

II - 1.2.1.4: Líquidos Iónicos Noutras Aplicações

Líquidos iónicos têm ganho importância crescente em outras inúmeras aplicações como é o caso da biocatálise, sendo usados em sistemas bifásicos com a água, reduzindo a inibição produzida pelo produto ou mesmo reagente se a reacção decorresse em ambiente exclusivamente aquoso. Os líquidos iónicos foram também utilizados em catálise enzimática na ausência total de água, nomeadamente lipases aumentando em alguns casos as percentagens de conversão e as enantioselectividades.³⁰

Os líquidos são usados também como fluidos em diversas aplicações de engenharia química tal como fluidos de transferência de calor usados em sistemas de refrigeração e sistemas de recolha de energia solar. Capacidades como a de armazenar grandes quantidades de calor, largo espectro de temperaturas em que lhe é permitido permanecer no estado líquido, baixa pressão de vapor e estabilidade térmica são propriedades que os convertem em fluidos de transferência de calor de eleição. O uso de líquidos iónicos como fluidos de engenharia não se esgota aqui, sendo capazes de favorecer a separação em misturas azeotrópicas, nomeadamente sistemas etanol/água, acetona/metanol, THF/água e metil-ciclo-hexano/tolueno. Possuindo uma viscosidade considerável, associada às muitas outras propriedades, os líquidos iónicos podem ainda ser usados como lubrificantes para superfícies aço/aço, aço/alumínio, aço/cobre entre muitas outras superfícies reduzindo extraordinariamente a fricção entre superfícies, sendo alguns deles superiores a lubrificantes actualmente no mercado.²⁶

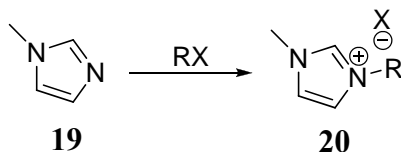
II - 1.2.1.5: Líquidos Iónicos Implementados na Indústria

Apesar do sucesso e ecletismo dos líquidos iónicos na substituição de produtos mais tradicionais numa grande variedade de aplicações, o seu uso ainda não se generalizou na indústria química e não só. Ainda assim alguns processos são realizados com a intervenção de líquidos iónicos, com melhorias significativas em relação aos processos anteriores. A título de exemplo a destacar o processo da BASF que em 1990 lançou o processo de fosgenação usando como catalisador sais de Vilsmeier, nomeadamente cloretos de (clorometileno)dialquilamónio, sais de baixo ponto de fusão²⁶. No mesmo ano, foi lançado pela IFP (Institut Français du Pétrole) o processo (DIFASOL) de dimerização do propeno em isómeros do hexeno na presença de catalisador de Níquel (II) e um líquido iónico.¹⁶ No ano de 2003, outro processo comercial foi lançado pela BASF, neste caso na reacção de formação de alcoxifenilfosfinas a partir fenil-clorofosfinas e álcoois. O *N*-metilimidazólio actua como base captando o protão proveniente do álcool, transformando-se num líquido iónico com ponto de fusão de 75°C, formando sistema bifásico facilmente separável do meio reaccional, é o denominado processo BASIL.³¹ As principais razões para a ainda escassa implementação de líquidos iónicos na indústria prendem-se com o seu ainda elevado custo, escasso conhecimento das suas propriedades físico químicas e presença de impurezas.

II - 1.2.2: Sais de Imidazólio

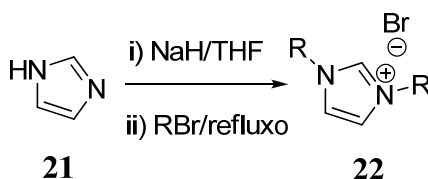
Líquidos iónicos baseados na unidade catiónica imidazólio são claramente os mais estudados tanto ao nível fundamental como ao nível de aplicações. Por outro lado a pureza do líquido iónico foi e é um assunto relevante consoante as propriedades físico-químicas que se estejam a estudar e as aplicações que se estejam a considerar. Tendo em conta os requisitos de pureza, custos associados e tempo reaccional foram desenvolvidos diferentes métodos de síntese que de certa forma são

complementares. Uma das primeiras metodologias de síntese disponível consistia na alquilação directa de um imidazólio **19** (**Esquema II.1**) (normalmente 1-metilimidazólio) através de um haloalcano, com bromoalcanos³² a proporcionarem tempos de reacção mais curtos que os correspondentes cloroalcanos.³³



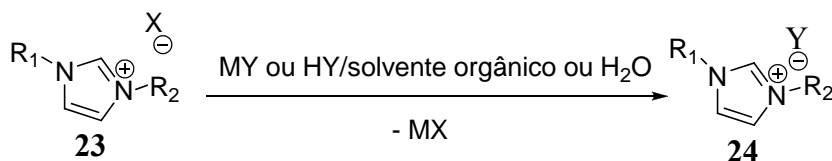
Esquema II.1: Alquilação do 1-metilimidazólio com um haleto de alquila, resultando num sal de imidazólio.

Com o objectivo de simplificar a síntese destes sais, dispensando o uso de solvente, e reduzir os tempos de reacção até 1 h., desenvolveu-se a síntese destes sais usando um microondas para o efeito.³⁴ Em todos estes métodos há um *work-up* de forma a isolar o produto ora através de extracção/lavagem e/ou recristalização. No caso do haloalcano ser gasoso à temperatura ambiente a reacção é realizada num autoclave a uma pressão de cerca de 6 atmosferas sobre N₂. No caso da preparação de sais de 1,3-dialquilimidazólio **22** (**Esquema II.2**), é colocado a reagir o imidazólio **21** na presença de hidreto de sódio e o bromoalcano pretendido.³⁵



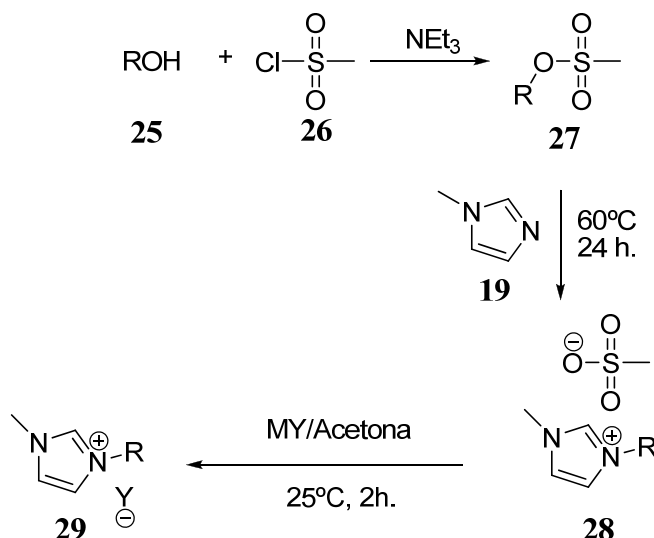
Esquema II.2: Síntese de um sal de 1,3-dialquilimidazólio.

Em todos estes casos é obtido um sal de imidazólio cujo anião é um haleto, normalmente cloreto ou brometo. Em muitos casos é importante que o líquido iónico seja líquido à temperatura ambiente ou que seja imiscível com a água ou até que contenha um anião não nucleófilo, nesse caso partindo do haleto de imidazólio (**20** ou **22**) é realizada a troca de anião na presença de um sal MY ou do ácido HY, em que M pode ser um catião metálico do Grupo 1 da tabela periódica (sódio, potássio e lítio os mais habituais) ou catião de prata e onde Y é o anião que se pretende trocar.



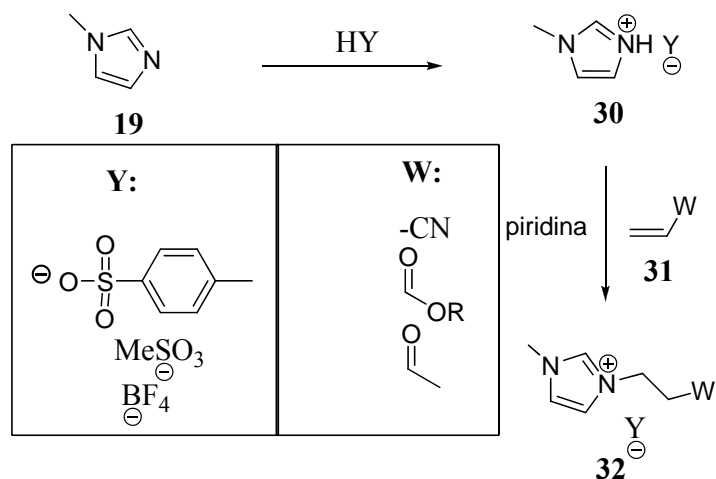
Esquema II.3: Esquema da troca de anião de um haleto X por outro tipo de anião Y.

O sal de prata é substancialmente mais dispendioso que sais de metais alcalinos ou de ácidos, mas também é aquele que permite a formação de haletos de prata mais facilmente separáveis do líquido iónico **24** obtido (**Esquema II.3**), o que na prática permite uma presença virtualmente nula de haleto como impureza em **24**. De forma a contornar a necessidade de usar o dispendioso sal de prata na reacção de troca de anião, foram desenvolvidos outros métodos de alquilação de imidazólio e, em alguns casos, conseguiram num só passo reacional o líquido iónico pretendido.



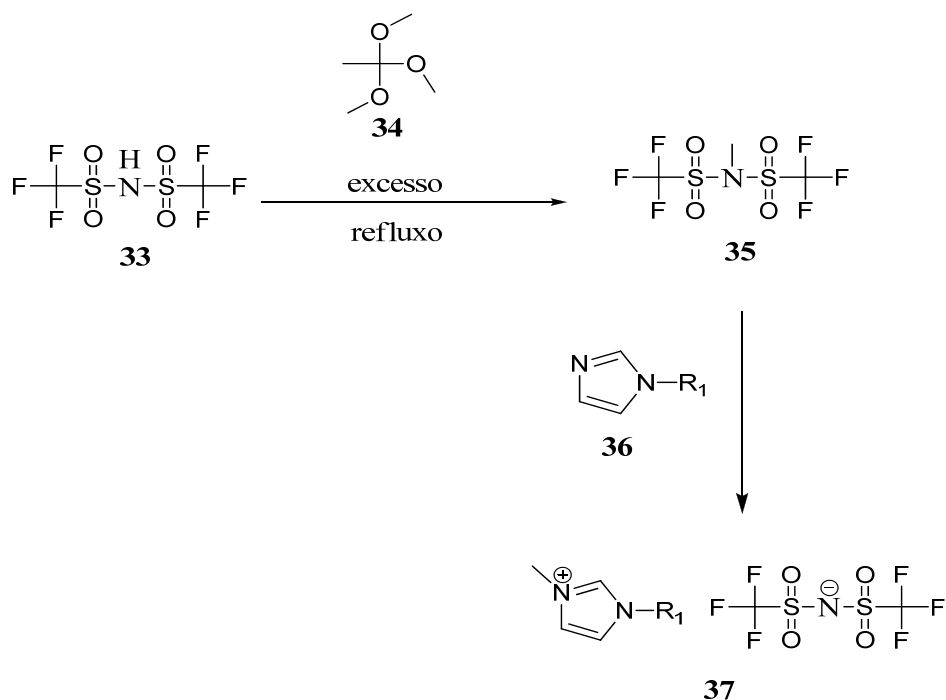
Esquema II.4: Esquema genérico de síntese de sais de imidazólio **28** e **29**.

Um método muito útil³⁶ consiste na alquilação de um imidazólio **19** na presença de um éster metanosulfónico **27** comercial, ou preparado a partir de um álcool **25** e do cloreto de metanosulfonilo **26** (**Esquema II.4**). O sal **28** assim preparado não possui contaminações de haletos, podendo facilmente ser trocado por outro anião para dar o líquido iónico desejado **29** usando na troca um sal de um metal alcalino MY.



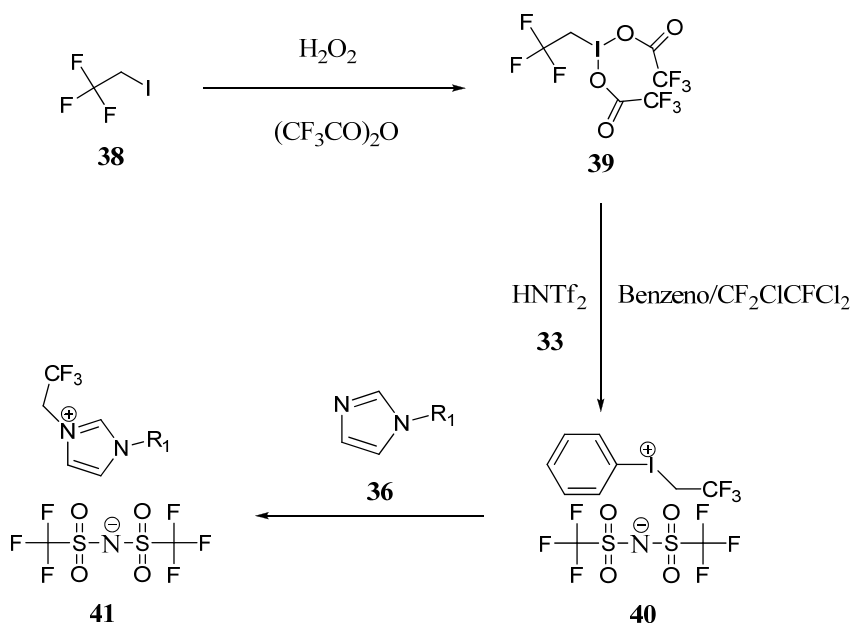
Esquema II.5: Esquema de síntese de um sal de imidazólio 32, baseado em adição de Michael.

Outro método, descrito por Wasserscheid *et al.*,³⁷ permite a obtenção directa de um líquido iónico com o anião pretendido sem passar por líquido iónico com anião halogeneto. O método consiste numa adição de Michael após reacção ácido base para introdução de anião. Este método de síntese permite a obtenção de líquidos iónicos funcionalizados com grupos electroatractores como nitrilos, esteres ou cetonas (**Esquema II.5**).



Esquema II.6: Síntese directa de sal bis(trifluorometilsulfonil)imidato 37 sem passar por halogeneto.

Tendo em conta que o anião bis(trifluorometilsulfonyl)imidato é um dos anião mais efectivos na obtenção de líquidos iônicos à temperatura ambiente, com elevada estabilidade térmica e baixas viscosidades, havia a necessidade da preparação directa destes sais sem passar antecipadamente por halogeneto correspondente. Nesse contexto surgiu o trabalho de Zhang *et al*³⁸, que usando excesso de ortoacetato de trimetilo **34** e bis(trifluorometilsulfonyl)imidato **33**, permitiram a obtenção de **35** que irá metilar imidazólio **36**, dando origem, num só passo de síntese, ao sal de imidazólio **37** com anião bis(trifluorometilsulfonyl)imidato (**Esquema II.6**).



Esquema II.7: Novo esquema de síntese de sal de imidazólio funcionalizado com cadeia fluorada **41, usando agente alquilante **40**.**

O mesmo autor³⁸ usa outro agente alquilante **40**, obtido através da oxidação de **38** com peróxido de hidrogénio e anidrido trifluoroacético e posterior reacção com benzeno e bis(trifluorometilsulfonyl)imidato **33**, para obtenção do sal de imidazólio fluoro-funcionalizado **41** (**Esquema II.7**). Apesar da variedade de sínteses disponíveis para obtenção de sais de imidazólio pode-se dizer que não há um método de síntese ideal relativamente aos parâmetros de pureza, simplicidade e robustez. a uma panóplia de reagentes mais abrangente, havendo sim métodos, alternativos, adaptados a situações específicas.

II - 1.2.3: Sais de Guanidínio

Os sais de guanidínio são uma classe de sais de catião orgânico que, apesar das semelhanças relativamente a outros sais de catiões orgânicos como os sais de imidazólio (possuindo baixa volatilidade e possibilidade de constituir sais líquidos à temperatura ambiente), possuem

características particulares como elevadas estabilidades, resultado da possibilidade de deslocalizar carga positiva do catião por quatro átomos distintos:

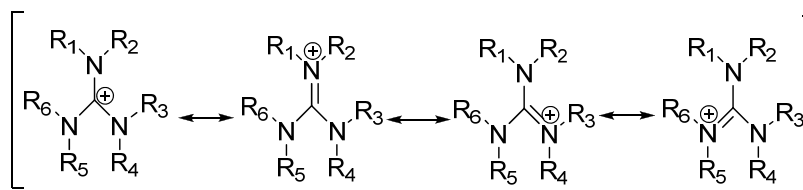


Figura II.5: As quatro possíveis estruturas de ressonância de um catião guanidínio.

Por outro lado os catiões guanidínio possuem uma geometria planar em Y, tal como seis posições (R_1 - R_6) com a possibilidade de estarem protonadas ou substituídas, sendo possível modelar a basicidade e interacção por pontes de hidrogénio com determinadas estruturas.

Mas não só a basicidade nem o número de pontos com capacidade de formar pontes de hidrogénio podem ser modelados pela natureza e número dos grupos substituintes, outras propriedades como o ponto de fusão, densidade, condutividade, viscosidade são influenciadas de forma significativa pela estrutura do sal de guanidínio. Formas mais simples da estrutura de um catião guanidínio são aquelas em que parte ou a totalidade dos seus grupos substituintes são protões sendo estas consideradas bases muito fortes (pK_a entre 12-13).³⁹ Na natureza a estrutura de guanidínio está presente no aminoácido arginina:

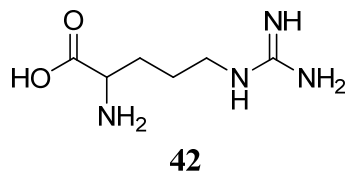


Figura II.6: Estrutura do aminoácido arginina.

Devido à capacidade de interacção por pontes de hidrogénio da estrutura de guanidínio, o aminoácido arginina é grandemente responsável pelo *folding* proteico, nomeadamente na formação de hélices α , mas não só, desempenha também papel relevante em centros activos enzimáticos como é o caso da carbopeptidase A, creatino cinase, fumarato redutase e malato desidrogenase. Esta estrutura tem também um papel relevante no controlo da informação, replicação e protecção do ADN através da interacções por pontes de hidrogénio entre o grupo guanidínio da arginina proteica e as cadeias fosfodiéster do DNA.³⁹ Tendo em conta estas propriedades únicas da unidade de guanidínio, sais baseados nesta unidade podem ter diversas aplicações, sendo talvez a mais imediata o seu uso em reconhecimento molecular, como é o caso de oxoaniões:

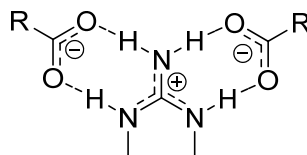
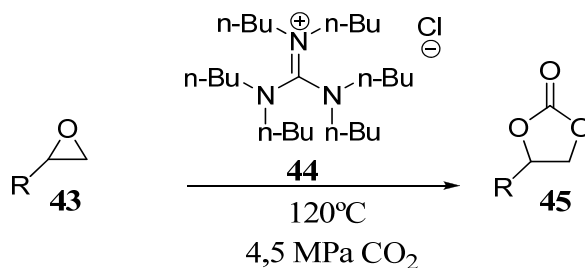


Figura II.7: Representação da peculiar interação entre unidade guanidínio e carboxilatos.

Mais concretamente, sais de guanidínio são usados no reconhecimento de fosfatos, fosfodiésteres, carboxilatos (**Figura II.7**), sulfatos e nitratos.³⁹ Propriedades como densidades moderadamente elevadas, baixas volatilidades fazem destes sais compostos com grande potencial para serem usados como compostos energéticos. Foram sintetizados nitratos e percloratos de guanidínios cíclicos e acíclicos apresentando alguns elevadas entalpias de formação,⁴⁰ sendo alguns destes compostos considerados líquidos iônicos. No campo dos materiais, foi descrito o uso de um sal de guanidínio, baseado no anião de triclanometano, como meio electrólito para a obtenção de células solares.⁴¹ Ao nível dos cristais líquidos, foram descritos sulfonatos de guanidínio,^{24,25} apresentando fases esmécticas.

Ao nível da síntese orgânica, sais tetra-alkuildimetilguanidínio foram referidos no ano de 2003 por Mateus *et al.*,⁴² como uma nova geração de líquidos iônicos possuindo características de estabilidade, solubilidade e densidade complementares às dos amplamente estudados sais de imidazólio. A partir daí a aplicação de sais de guanidínio como meio reaccional e catalisador em reacções várias não parou de aumentar, constituindo uma mais valia em termos de rendimento, velocidade de reacção, selectividade, já para não falar no aspecto da sustentabilidade. Em 2004, Xie *et al.*⁴³ usaram um guanidínio cíclico tanto como meio reaccional como catalisador de transferência de fase na oxidação selectiva de álcoois benzílicos em sistema bifásico (H₂O/LI), usando hipoclorito de sódio como agente oxidante. Os rendimentos e selectividades foram consideravelmente superiores aos métodos utilizando solvente orgânico volátil e catalisador de transferência de fase ou usando exclusivamente sal de imidazólio. Em 2005, Xie *et al.*⁴⁴ descreveram o uso de cloretos de hexabutilguanidínio em reacção de fixação de CO₂ por parte de epóxidos (**Esquema II.8**).

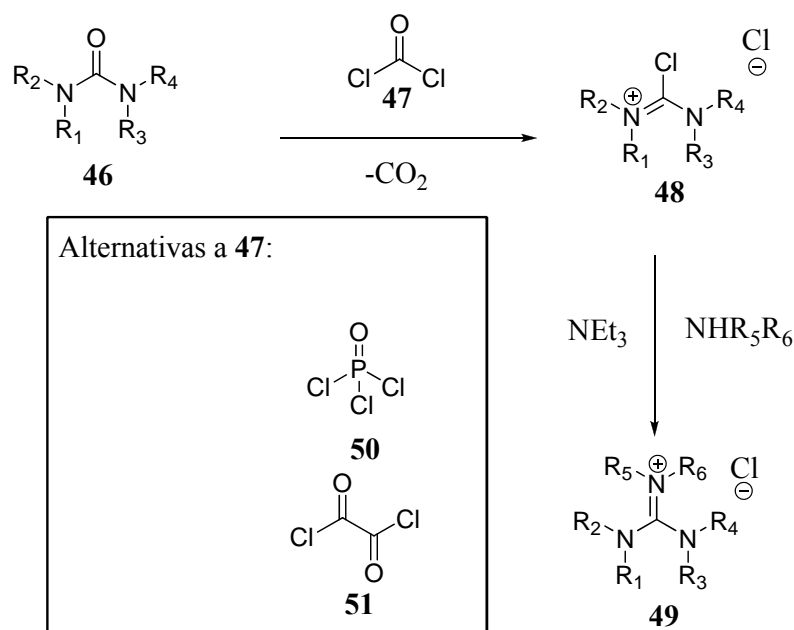


Esquema II.8: Reacção de fixação de CO₂ por parte de epóxidos 43, catalisada pelo sal de guanidínio 44.

Neste caso o sal de guanidínio participa nesta reacção como catalisador na abertura do epóxido, abrindo este pelo lado menos substituído. A reacção processa-se com elevada selectividade, rendimentos muito elevados, e tempos de reacção curtos podendo ser reciclado 5 vezes sem perda significativa de actividade catalítica, não sendo necessário utilizar solvente.

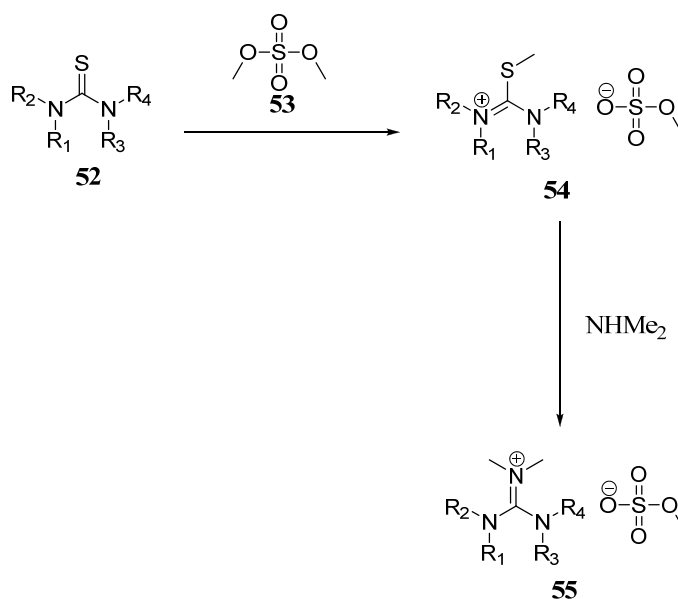
Em 2006, Li *et al*⁴⁵ usaram um guanidínio, penta-alquilado, em reacção de Heck, actuando como meio reaccional polar, como base e ligando de Pd(II), resultando em rendimentos elevados, tempos de reacção curtos, selectividades elevadas, sendo ainda um meio que é possível reutilizar até 5 vezes, conduzindo ainda a uma redução do Paládio utilizado (até 0,01 mol %) sem perda significativa de actividade catalítica. Ainda em 2006, Zhu *et al*⁴⁶ usou sais de tetrametilguanidínio como catalisadores em reacção de aldol de aldeído aromático e cetona assimétrica, activando, na forma de enamina, cetona assimétrica, favorecendo rendimento e selectividade de ataque nucleofílico ao aldeído pelo lado mais substituído da cetona. Já em 2007, Xin *et al*⁴⁷ usaram sal de guanidínio cíclico como meio reaccional em condensação de Knoevenagel, sendo um meio robusto, conduzindo a elevados rendimentos e tempos de reacção extremamente curtos para uma variedade considerável de substratos. A possibilidade de reciclar este meio até seis vezes sem perda significativa de actividade catalítica é uma mais valia adicional. Sais de guanidínio foram ainda utilizados em substituição nucleofílica de haletos de alquilo, reacção de Mannich, reacção de Henry,⁴⁷ hidrogenação, hidroformilação⁴⁸ e dihidroxilação de Sharpless assimétrica,⁴⁹ sendo ainda utilizados como meios indutores de quiralidade em reacções inserção assimétrica de C-H catalisada por Rh(II) para formação γ -lactama e dihidroxilação de Sharpless assimétrica, com valores de excesso enantiomérico para ambas as reacções⁵⁰ comparáveis em relação a valores de referência.

Relativamente à síntese de sais de guanidínio estão referenciados muitos métodos complementares com complexidade variável, associada ao número de substituintes distintos que se pretende que a unidade guanidínio tenha. Em 1984, Kantlehner *et al*⁵¹ propõe método de síntese partindo de uma ureia **46** (**Esquema II.9**):



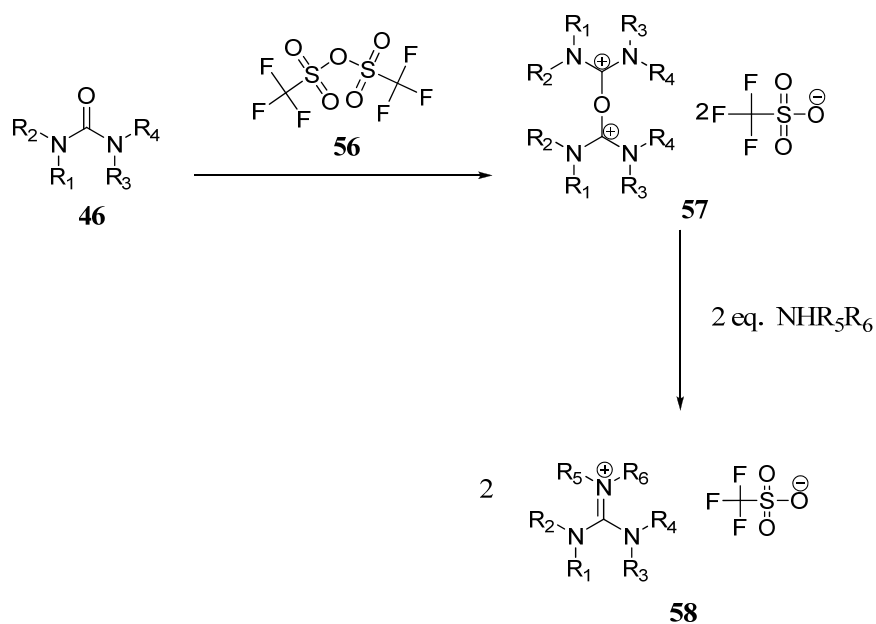
Esquema II.9: Método de síntese de sal de guanidínio 49 a partir de ureia 46.

No primeiro passo desta síntese a ureia **46** reage com fosgênio **47** resultando o sal **48** que na presença de uma amina secundária e base trietilamina irá resultar o sal de guanidínio **49**, este método de síntese permite a obtenção de uma grande variedade de sais de guanidínio, tendo em conta a possibilidade de variar os grupos R_1 , R_2 , R_3 , R_4 da ureia **46** e R_5 , R_6 na amina secundária utilizada. Existem variações deste método usando no primeiro passo de síntese o oxiclreto de fósforo⁴³ **50** ou cloreto de oxalilo⁵² **51**. Noutro método⁵² de síntese parte-se de uma tioureia **52**, que é metilada na presença de dimetilsulfato **53** resultando no sal **54**, que na presença de dimetilamina irá resultar o sal de guanidínio **55** no segundo passo reaccional (**Esquema II.10**).

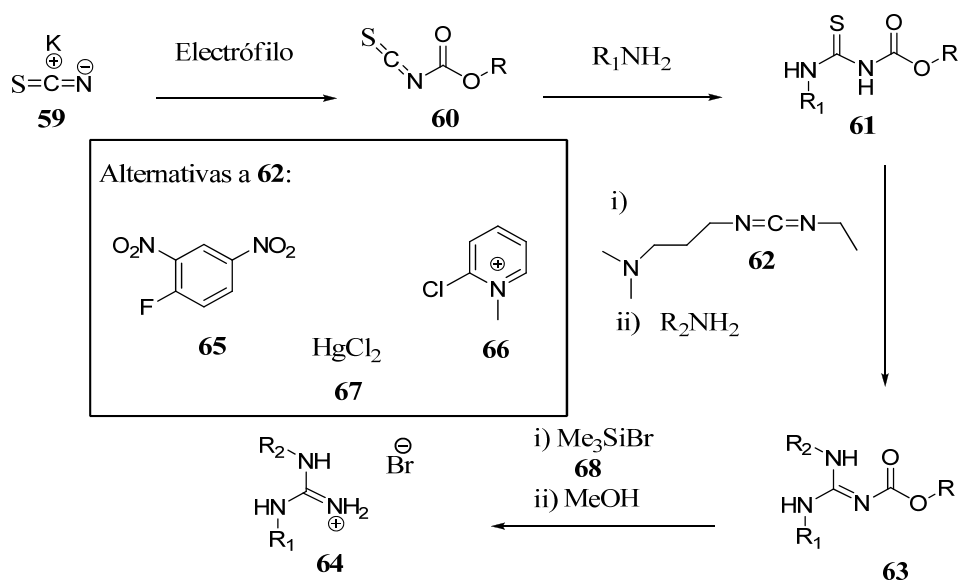


Esquema II.10: Esquema de síntese de um sal de guanidínio 55 a partir de uma tioureia 52.

Variação deste método⁵² consiste na utilização de uma amina primária que ao reagir com os sais **48** ou **54** irá resultar numa guanidina pentassubstituída que pode posteriormente ser alquilada por um haleto de alquilo ou outro agente alquilante, havendo assim um maior potencial na obtenção de guanidínio com maior variabilidade de cadeias R_1 - R_6 . Um novo método de síntese⁵² consiste no uso de anidrido triflico **56** que reage com ureias **46**, resultando no sal **57**, que na presença de amina secundária resultará no triflato de guanidínio **58** no segundo passo reaccional (**Esquema II.11**).

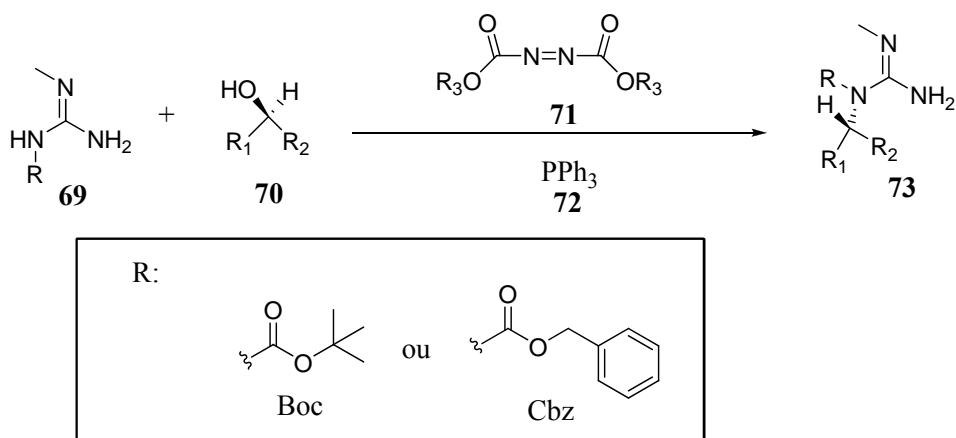


Esquema II.11: Esquema de síntese de triflato de guanidínio 58, usando anidrido triflico 56 e amina secundária.



Esquema II.12: Esquema de síntese de um sal de guanidínio 64 a partir do tiocianato de potássio 59.

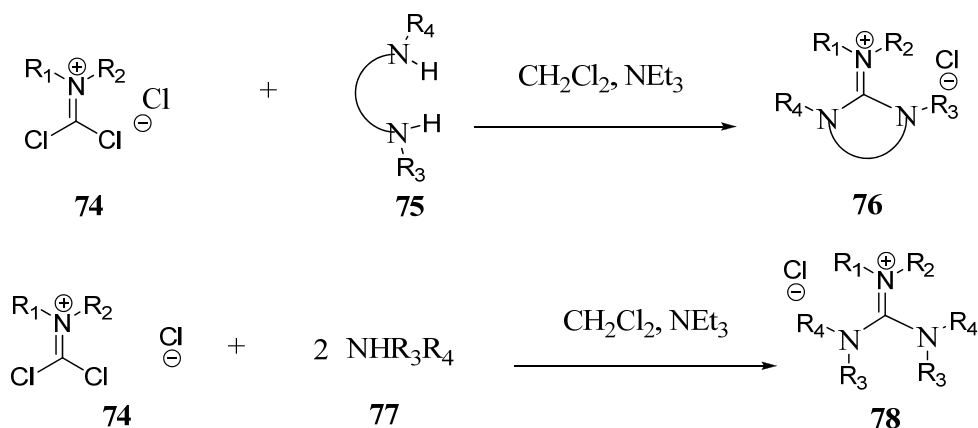
Noutro método de síntese,⁵³ partindo do tiocianato de potássio, um substrato muito simples, é possível obter um sal de guanidínio **64** (**Esquema II.12**), para tal o tiocianato de potássio é activado na presença de um electrófilo para dar o composto **60**, que no segundo passo reaccional, na presença de amina primária, irá resultar a tioureia **61**. Esta tioureia depois de ser activada pela, 1-(3-dimetilaminopropil)-3-etil-carbodiimida **62** é colocada a reagir com amina primária, resultando na guanidina protegida **63**, tudo isto no terceiro passo reaccional. No quarto passo reaccional a guanidina **63** é desprotegida na presença de brometo de trimetilsililo **68** e metanol para dar origem ao brometo de guanidínio **64**. No terceiro passo reaccional, em vez de usar a carbodiimida **62**, é possível usar como agente activador o reagente de Sanger⁵⁴ **65**, o reagente de Mukayama **66**,⁵⁵ ou o cloreto de mercúrio **67**.⁵⁶ Um outro método descrito,⁵⁷ consiste na alquilação de uma guanidina protegida com Boc ou Cbz. Essa alquilação é feita com um álcool, nomeadamente quiral, segundo o protocolo de Mitsunobu. Nesta síntese o álcool é activado através do composto azo **71** e a trifenilfosfina **72**, sendo assim usado como agente alquilante, esta é uma reacção muito selectiva para alquilação no azoto protegido com Boc ou Cbz para a obtenção da guanidina **73** (**Esquema II.13**), que facilmente pode ser protonada ou alquilada para obtenção do correspondente sal de guanidínio. Esta é uma forma de introduzir quiralidade num guanidínio.



Esquema II.13: Esquema de alquilação de uma guanidina protegida **69 com um álcool **70** usando o protocolo de Mitsunobu.**

Um método bastante atractivo,^{42,58} pela sua simplicidade (um passo reaccional), e pela possibilidade de, a partir de reagentes relativamente simples poder preparar uma grande variedade de sais de guanidínio, consiste na reacção de sais de *N,N*-dialquilfosgeniminio **74**, na presença de diamina **75** ou 2 equivalentes de uma amina secundária **77**. Desta reacção resultam os cloretos de guanidínio **76** e **78** (**Esquema II.14**). Os cloretos de *N,N*-dialquilfosgeniminio **74**, nomeadamente o cloreto de *N,N*-dimetilfosgeniminio, são muito reactivos sendo manipulados a baixas temperaturas e

sob atmosfera de azoto. Este foi o método de síntese utilizado ao longo deste trabalho para obtenção de novos sais de guanidínio com propriedades físico-químicas pretendidas.



Esquema II.14: Método simples de síntese de sais de guanidínio 76 e 78 a partir de cloreto de *N,N*-dialquilfosgeniminio 74.

Muitos dos sais obtidos por este método de síntese são cloretos ou brometos. Para a obtenção de uma maior variedade de sais de guanidínio, normalmente procede-se à troca de anião tal como descrito anteriormente para sais de imidazólio, usando para tal sais de metais alcalinos, sais de prata ou ácidos.

Depois de descritos os principais métodos de síntese de sais de guanidínio é de referir ainda uma característica particular destes sais, a ter em conta na hora de identificar o guanidínio sintetizado, nomeadamente em espectros de ¹H-RMN. No caso das cadeias substituintes, ligadas ao mesmo átomo de azoto da unidade guanidínio, serem diferentes, há a possibilidade de isomerismo E/Z à volta da ligação dupla parcial do imínio (**Figura II.8**).

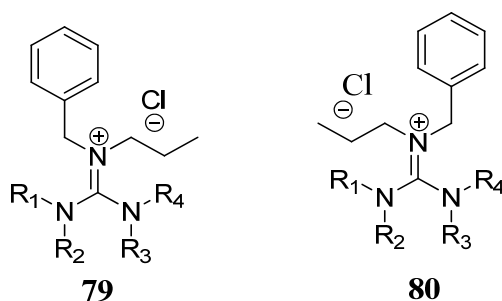


Figura II.8: Isomeria E/Z num sal de guanidínio.

Esta observação é válida se R₁, R₂ são diferentes de R₃, R₄. Neste caso o espectro de ¹H-RMN poderá apresentar diferentes desvios para os diferentes isómeros.⁵²

II - 1.2.4: Obtenção de Líquidos Iónicos Purificados

Uma das principais razões para ainda escassa implementação dos líquidos iónicos na indústria e outras áreas de actividade é a presença de impurezas, ora provenientes dos reagentes utilizados, ora produzidos durante o processo de síntese, ora obtidos pela própria natureza do líquido iónico, nomeadamente o facto de alguns serem higroscópicos. Cloretos e água são as impurezas mais comuns. Relativamente aos efeitos da presença de cloretos, estes coordenam com metais de transição utilizados em catálise tendo impacto normalmente negativo na velocidade de reacção, relativamente a propriedades físico-químicas, a presença de cloretos normalmente provoca um incremento significativo na viscosidade e um decréscimo na densidade. Já a água, em geral, pode provocar uma diminuição da viscosidade tal como um decréscimo na densidade.⁵⁹ Outras impurezas, que nomeadamente podem dar alguma coloração (e não só) aos líquidos iónicos poderão ser relevantes, principalmente para uso espectroscópico.⁶⁰

Há procedimentos a realizar, antes, durante e depois da síntese de líquidos iónicos para evitar a presença de impurezas. Antes da síntese pode-se destilar reagentes e solventes na presença de fio de sódio, hidreto de cálcio ou outro agente que efectivamente desidrate. Neste caso está-se a evitar a presença de impurezas dissolvidas, outros solventes e água. Durante a síntese, deve-se ter atenção na adição de reagentes, normalmente adições lentas a temperaturas baixas conduzem a produtos com menos impurezas. As reacções devem ser realizadas sob atmosfera inerte, nomeadamente azoto, podendo optar-se por proteger o meio reaccional da luz envolvendo o balão em folha de alumínio. Após a reacção, para além dos procedimentos habituais de *work-up* (extracções, lavagens, evaporação de solventes) poderão ainda ser tomadas diversas medidas para purificação, no caso do líquido iónico ser sólido à temperatura ambiente pode-se tentar recrystalizá-lo; outro método passa pelo uso de carvão activado, especialmente eficaz na eliminação de impurezas que dão coloração ao LI. Pode-se ainda passar o LI por uma coluna de sílica, muito embora se corra o risco de ter algumas partículas de sílica dissolvida no líquido iónico.⁶⁰ Posteriormente, deve-se colocar o LI sob pressão reduzida e temperatura de forma a remover solventes orgânicos e água usados na síntese e *work-up*. LIs muito higroscópicos acarretam dificuldades acrescidas, havendo alguns que virtualmente seriam sólidos, mas na prática são líquidos pela dificuldade em eliminar moléculas de água associadas à estrutura do LI. Uma forma alternativa de purificar LIs, é através da sua destilação pressão reduzida,⁶¹ algo até há pouco tempo impensável dado o rótulo de não volatilidade atribuído desde cedo aos líquidos iónicos. Na conservação de LIs deverá evitar-se locais húmidos, luz e calor.

II-1.3: Propriedades Físico-Químicas de Líquidos Iônicos

II - 1.3.1: Ponto de Fusão

O ponto de fusão (*p.f.*) é uma propriedade fundamental dos compostos químicos, consiste na temperatura ou gama de temperaturas em que ocorre mudança de fase do estado sólido para o estado líquido, isto estando a falar de aumento de temperatura. No sentido inverso o ponto de solidificação (*p.s.*) marca a passagem do estado líquido para o estado sólido, e em teoria deveria corresponder à temperatura, ou gama de temperaturas em que ocorre o ponto de fusão, mas na prática há pequenas diferenças. Outra característica do ponto de fusão é que se aplica, em geral, a sólidos cristalinos, ou seja, a compostos que têm um arranjo ordenado repetitivo constituindo uma rede cristalina. Mas nem todos os sólidos são cristalinos nem a passagem para o estado líquido significa perda de cristalinidade, no primeiro caso os sólidos podem ser amorfos, não apresentando ordem ao nível do posicionamento atômico, pelo menos numa gama alta de distâncias, como é o caso de muitos polímeros, vidros baseados em silício, ou até mesmo compostos com tendência para formar cristais, mas que em condições de arrefecimento muito rápido também formam sólidos amorfos. A passagem de um estado sólido amorfo para o estado líquido define a transição vítrea (*t.v.*). O outro caso refere-se à existência de cristais líquidos, em que a fusão não marca a perda de cristalinidade, havendo associadas outras transições em função de temperatura que marcam a perda de cristalinidade ou de um tipo de cristalinidade. Também poderá acontecer um determinado composto apresentar tanto uma transição vítrea como um ponto de fusão, ou mesmo dois pontos de fusão distintos, facto que pode acontecer quando existem arranjos distintos inerentes ao mesmo composto (polimorfos), o que dificulta ainda mais a definição da passagem do estado sólido para o estado líquido. Em síntese, a passagem do estado sólido para o estado líquido não é uniforme para todos os compostos havendo oscilação no significado físico-químico que essa transição pode ter, mesmo focando a atenção apenas nos casos que apresentem ponto de fusão; todos estes factos contribuem e muito para a complexidade desta propriedade físico-química. Para além deste factor, esta propriedade físico química caracteriza-se por reflectir tanto o arranjo dos átomos na rede cristalina como as interacções inter- e intramoleculares, características interdependentes (não isoláveis) o que contribui ainda mais para a complexidade do ponto de fusão. O grau de complexidade do ponto de fusão é comparável com a sua importância senão veja-se a panóplia de processos fundamentais ou aplicações em que o ponto de fusão está envolvido, em primeiro lugar define a temperatura a partir da qual um determinado composto está disponível no estado líquido regendo a aplicabilidade dum determinado composto como meio reaccional, meio energético, biológico ou até mesmo como meio de transporte. O ponto de fusão também é útil na identificação e caracterização de um determinado composto, ou na obtenção de outras propriedades físico-químicas com que o ponto de fusão se correlaciona como é o caso da solubilidade de um composto num determinado solvente, mais concretamente a solubilidade em água⁶² definida como:

$$\log S_m = -\frac{\Delta S_f(p.f. - 25)}{1364} - \log P + 0,8$$

Equação II.1

em que ΔS_f é a entropia de fusão, $p.f.$ o ponto de fusão e $\log P$ é o logaritmo do coeficiente de partição entre octanol e água. Saber a solubilidade em água é muito importante visto a síntese de novos fármacos estar em grande medida dependente da sua solubilidade dependendo de se querer que o fármaco tenha um efeito mais imediato e mais concentrado num curto espaço de tempo, ou se deve ter uma actividade durante um período de tempo mais prolongado, já para não falar na avaliação do potencial tóxico de um determinado composto. Devido ao enorme interesse em saber antecipadamente o ponto de fusão de um determinado composto, e apesar da complexidade inerente a esta propriedade, foram encontradas relações, criados modelos e desenvolvido metodologias mais ou menos gerais, lineares ou não lineares, e que visam a previsão/interpretação do ponto de fusão a partir de características estruturais e/ou outras propriedades físico-químicas desse composto.

A grande maioria destes modelos parte do pressuposto de que à temperatura de fusão $p.f.$ (K):

$$p.f. = \frac{\Delta H_f}{\Delta S_f}$$

Equação II.2

em que ΔH_f é a entalpia de fusão (kJ.mol^{-1}) dependente em grande medida das forças inter-moleculares, ΔS_f (kJ.K^{-1}) é entropia de fusão, relacionada com factores geométricos inerentes à própria molécula.

A **Equação II.2** deriva do facto do processo de fusão, a um gradiente de temperatura suficientemente diminuto, ser reversível, ou seja, a energia ganha na fusão é igual à energia perdida no processo inverso, nessas condições:

$$\partial S = \frac{\partial q_{rev}}{T}$$

Equação II.3

em que ∂S é a variação infinitesimal da entropia, ∂q_{rev} é diferencial inexacto (depende do caminho percorrido) do calor envolvido num processo reversível.

Considerando o sistema (onde se dá a fusão) e o exterior (fonte do calor responsável pela fusão):

$$\partial S_{Total} = \partial S_{sistema} + \partial S_{exterior}$$

Equação II.4

Num processo reversível, em que o calor ganho pelo sistema é igual ao calor perdido pelo exterior (mas de sinal inverso), resulta:

$$\partial S_{Total} = 0 = \partial S_{sistema} - \frac{\partial q_{rev}}{T_{exterior}}$$

Equação II.5

Rearranjando a **Equação II.5**:

$$\partial S_{sistema} = \frac{\partial q_{rev}}{T_{exterior}}$$

Equação II.6

Num sistema isotérmico, como é o caso do sistema onde está a decorrer uma fusão, a integração da **Equação II.6** resulta:

$$\Delta S_{sistema} = \frac{q_{rev}}{T_{exterior}}$$

Equação II.7

No caso de pressão constante, o calor envolvido é a entalpia:

$$\Delta S = \frac{\Delta H}{T}$$

Equação II.8

Rearranjando a **Equação II.8** e substituindo T pelo ponto de fusão resultará a **Equação II.2**.⁶³ Pela equação é possível observar que o ponto de fusão depende tanto de factores entálpicos como entrópicos, os primeiros relacionados com as forças inter-moleculares que regem o estado sólido, os segundos relacionados com factores geométricos das próprias moléculas. Conjugando factores tanto entálpicos como entrópicos foram desenvolvidos modelos mais ou menos gerais que visavam a previsão de pontos de fusão e a sua interpretação segundo características estruturais da própria molécula.

No caso de moléculas simples como é o caso dos alcanos a relação entre a estrutura e o ponto de fusão é praticamente linear pois começando no metano, que tem um ponto de fusão de -182,5°C, aumentando progressivamente o tamanho da cadeia o ponto de fusão também vai aumentando, o icosano já possui um ponto de fusão de 36,8°C, essa tendência é devida ao facto de, ao aumentar a

cadeia, as interações de Van der Waals aumentarem, aumentando consequentemente o ponto de fusão. A essa tendência só há a acrescentar dois factos, primeiro até ao propano o ponto de fusão decresce, ainda que seja pouco, isso deve-se ao facto de no estado sólido (estado condensado da matéria) para além das forças inter-moleculares, também é relevante a geometria da molécula e a forma como se acomodam as moléculas umas nas outras, e o metano, quase esférico é mais fácil de acomodar que o etano ou o propano. Em segundo lugar, aumentando a cadeia passando de um número ímpar para um número par de átomos de carbono o ponto de fusão aumenta muito mais pronunciadamente do que passando de um número par para um número ímpar de átomos, esta observação deve-se ao facto de que as moléculas com número par de átomos se acomodarem melhor na rede cristalina do que as moléculas de número ímpar. Já nos cicloalcanos a tendência geral é também de aumentar o ponto de fusão com o aumento do número de átomos de carbono do anel, mas neste caso os cicloalcanos de número ímpar de átomos vão ter ponto de fusão mais baixo que o cicloalcano de número de átomos par anterior. Outro facto a salientar é que linhas gerais os cicloalcanos vão ter pontos de fusão mais elevados que os seus homólogos lineares, isto deve-se ao facto de nos compostos cíclicos os graus de liberdade conformacionais são menores, diminuindo a entropia do sistema e consequentemente aumentando a eficácia em acomodar as moléculas na rede cristalina. Esta interpretação directa do ponto de fusão em função da estrutura do composto é acessível para moléculas simples⁶⁴ como as que foram focadas acima, no caso de moléculas mais complexas, com mais funcionalidades distintas, a interpretação/obtenção de pontos de fusão recorrendo a modelos matemáticos/computacionais torna-se fundamental. Zhao e Yalkowsky⁶⁵ obtiveram um modelo, usando para tal 1040 compostos alifáticos possuindo diversas funcionalidades. Para tal partiram da relação expressa na **equação II.2**, em que a entalpia de fusão é função da contribuição individual de grupos de átomos presentes em cada molécula (46 grupos distintos, desde contribuição de grupos como “-CH₂-“ ou presença de halogénios, em adição a 9 factores de correcção, relacionados com a presença de anéis ou no caso da molécula ser zwitteriónica).

A entalpia de fusão ficou assim definida:

$$\Delta H_f = \sum n_i . m_i$$

Equação II.9

em que n_i é a ocorrência do grupo i em determinado composto e m_i a contribuição que o grupo i vai ter na entalpia de fusão. Cada valor de m_i foi determinado empiricamente para o conjunto de 1040 moléculas alifáticas.

Relativamente à contribuição entrópica, esta representa as modificações dos graus de liberdade translacionais, rotacionais e conformacionais envolvidas na mudança de fase.

$$\Delta S_f = \Delta S_f^{trans} + \Delta S_f^{rot} + \Delta S_f^{conf}$$

Equação II.10

A contribuição translacional tem peso negligenciável visto que, regra geral, nas fusões as variações do volume são pequenas (10-15%) permitindo mais liberdade apenas ao nível rotacional e conformacional. Assim, contabilizando apenas estas duas contribuições irá resultar:

$$\Delta S_f = 56,5 - 19,2 \log \sigma + 9,2\tau$$

Equação II.11

Resultado da modificação⁶⁶ da regra de Walden segundo a qual a entropia para compostos rígidos como estruturas cíclicas ou aromáticas tem um valor constante de entropia de fusão de 56,5 J K⁻¹ mol⁻¹. Em que σ é a simetria rotacional, e indica o número de imagens idênticas resultado da rotação rígida da molécula, e τ é a flexibilidade molecular e indica o número de ângulos torsionais de uma molécula e é definido por:

$$\tau = SP_3 + 0,5.SP_2 + 0,5.ANEL - 1$$

Equação II.12

Neste caso SP_3 é o número de átomos sp^3 , que não pertençam a um anel nem sejam átomos terminais. SP_2 é o número de átomos sp^2 que não pertençam a um anel nem sejam átomos terminais. $ANEL$ é o número de sistemas cíclicos na molécula. Substituindo as **Equações II.9 e II.11** na **Equação II.2** irá resultar:

$$p.f. = \frac{\Delta H_f}{\Delta S_f} = \frac{\sum n_i.m_i}{56,5 - 19,2 \log \sigma + 9,2\tau}$$

Equação II.13

Esta equação permite explicar o ponto de fusão de 1040 compostos alifáticos usando unicamente parâmetros relacionados com a estrutura molecular, com um RMS (raiz do erro médio quadrático) de 34,4 K.

Na mesma linha, baseado na relação termodinâmica da **Equação II.2**, Abramowitz e Yalkowsky⁶² obtiveram um modelo que visa a obtenção de pontos de fusão, baseado num conjunto de 85 compostos aromáticos com grupos substituintes que não formassem pontes de hidrogénio a equação obtida foi a seguinte:

$$p.f. = 0.772.p.e. + 110,8.SIGMAL + 11,56.ORTO + 31,9.EXPAN - 240,7$$

Equação II.14

Em que *p.e.* é o ponto de ebulição do composto (obtido experimentalmente mas que se pode calcular usando propriedades aditivas dos diversos grupos presentes numa molécula), e representa o parâmetro entálpico do ponto de fusão. Os parâmetros entrópicos relacionados com a geometria da molécula e presentes na **Equação II.14** são *SIGMAL*, definido pelo logaritmo da simetria rotacional σ , *ORTO*, que indica o número de grupos que estão orto em relação a outros substituintes aromáticos (indicativo das tensões estéricas), *EXPAN*, excentricidade elevada ao cubo, em que a excentricidade (quantifica o desvio em relação à forma esférica das moléculas) está definida por:

$$Excentricidade = \frac{\text{comprimento}_{\text{molecular}}_{\text{máximo}}}{\text{diâmetro}_{\text{molecular}}_{\text{médio}}}$$

Equação II.15

Na **Equação II.14**, assume-se que as interacções responsáveis pelo ponto de ebulição são similares às interacções responsáveis pelo ponto de fusão, só que no estado sólido as distâncias intermoleculares são menores, logo as interacções intermoleculares envolvidas na ebulição e fusão não podem ser interpretadas da mesma maneira daí a presença dos outros parâmetros na **Equação II.14**. Este modelo apresenta um RMS de 22,8°C para o conjunto de 85 compostos aromáticos. Os dois modelos anteriores ora se aplicavam exclusivamente a compostos alifáticos ou apenas a compostos aromáticos. Mais recentemente foi construído um modelo mais geral,⁶⁷ incluindo moléculas alifáticas e aromáticas recorrendo a redes neuronais de retropropagação (back propagation-neural networks, BPG-NN) método de aprendizagem não linear. Este modelo baseado em 4173 compostos foi construído usando análise de componentes principais (principal component analysis, PCA) de forma a reduzir a dimensão do espaço vectorial ocupado por descritores moleculares 2D (obtidos a partir da conectividade da molécula sem ter em conta a conformação 3D da molécula). Ao nível de resultados foi obtido RMS de 48°C (R^2 de 0,661) para conjunto de treino de 2089 objectos, 49,3°C para conjunto de teste de 1042 objectos, 50,4°C para conjunto interno de validação de 1042 objectos e 41,4°C para conjunto externo de validação de 277 objectos. O valor de RMS é maior usando este modelo geral do que nos dois modelos anteriores (comparando os conjuntos de treino). Este modelo tem a desvantagem de ser menos interpretável que os anteriores, ainda assim assinala uma forte relação entre o ponto de fusão e o tamanho e polaridade das moléculas.

II - 1.3.1.1: Ponto de Fusão de Líquidos Iónicos

Até agora, foi caracterizado o ponto de fusão, dando particular atenção a moléculas orgânicas não carregadas, como foi o caso dos três modelos anteriormente focados. Irá ser focado agora o caso particular dos líquidos iónicos, objectos de estudo neste capítulo. Os líquidos iónicos tal como o nome indica estão constituídos por iões, catião e anião em que pelo menos um destes é orgânico (normalmente o catião). Trata-se portanto de sais que, contrariamente à generalidade dos sais inorgânicos (como o caso do NaCl), fundem a uma temperatura baixa, próxima da temperatura ambiente. Para além de factores geométricos, que condicionam o acomodar das estruturas na rede cristalina, outra razão fundamental para o facto dos líquidos iónicos terem um ponto de fusão baixo prende-se com a combinação de interacções inter-moleculares características dos líquidos iónicos e que incluem, para além das interacções iónicas (típicas de sais inorgânicos), outro tipo de interacções (típicas de moléculas neutras, como a generalidade das moléculas orgânicas) tais como forças de Van der Waals (que incluem interacções dipolo-dipolo e forças dispersivas de London), interacções por pontes de hidrogénio, interacções π - π e interacções entre electrões não ligantes e orbitais π .⁶⁸ Esta panóplia de interacções possíveis faz dos líquidos iónicos compostos extremamente complexos mas ao mesmo tempo com uma gama impressionante de possibilidades de aplicações como meio reaccional, podendo adaptar-se às condições particulares de uma dada reacção segundo a sua fórmula estrutural, entre outras aplicações, daí a curiosidade e interesse em estudar as propriedades dos líquidos iónicos, em particular o ponto de fusão, que marca a passagem para o estado líquido e o consequente uso dos líquidos iónicos como meio reaccional. A equação termodinâmica que rege o ponto de fusão dos líquidos iónicos é a mesma descrita acima para a generalidade dos casos (**Equação II.2**), o parâmetro entrópico desta equação é composto pela entropia associada aos graus de liberdade, translacionais, rotacionais e conformacionais, como está descrito na **Equação II.10**. O parâmetro entálpico responsável pela fusão por sua vez está definido por:

$$\Delta H_f = \Delta H_f^{ionico} + \Delta H_f^{Van} + \Delta H_f^{SC}$$

Equação II.16

em que ΔH_f^{ionico} é a componente entálpica das interacções iónicas, ΔH_f^{Van} é o parâmetro entálpico de interacções de Van der Waals e ΔH_f^{SC} é a entalpia que caracteriza as interacções por pontes de hidrogénio, interacções π - π e interacções n- π . Partindo das **equações II.2, II.10 e II.16** e partindo de observações feitas por alguns autores^{62,65,69,70,71} válidas principalmente para a estrutura do catião (geralmente mais complexo que o anião) é possível associar alguns traços estruturais ao comportamento termodinâmico do composto relativamente ao ponto de fusão.

Geralmente estruturas lineares permitem pontos de fusão mais baixos que equivalentes cíclicos, pelo maior número de graus de liberdade conformacionais associados às estruturas lineares (entropia conformacional mais elevada). Capacidade de deslocalização da carga permite pontos de fusão mais baixos, pela diminuição do factor entálpico. Relativamente à assimetria essa também diminui o ponto de fusão, principalmente pelo aumento dos graus de liberdade rotacionais associados à entropia. Aumentando as ramificação da molécula a tendência geral é de aumentar o ponto de fusão, visto as conformações à volta da ramificação estarem mais constrangidas, reduzindo o factor da entropia conformacional, aumentando o ponto de fusão. No caso do tamanho das cadeias alquílicas, aumentando-as vai ser reduzido o ponto de fusão até certo ponto, principalmente por factores de entropia conformacional dominarem sobre factores entálpicos. Continuando a aumentar o tamanho da cadeia alquílica, a partir de certa altura as forças de Van der Waals já começam a ter um grande peso e o aumento da cadeia irá resultar em incremento no ponto de fusão. Ou seja o ponto de fusão vai ter um comportamento em “U” com o aumento do tamanho da cadeia alquílica (especialmente no caso de sais de 1-alquil-3-metilimidazólio). A propósito da alteração do factor entálpico de fusão associado ao aumento do tamanho das cadeias alquílicas, Santos *et al.*,⁷² recorrendo a cálculos de dinâmica molecular determinaram a entalpia de vaporização (confirmada experimentalmente) para a série 1-alquil-3-metilimidazólios de bis(trifluorometilsufonil)imidatos; para o cálculo da entalpia de vaporização são determinadas as energias coesivas (ΔU) associadas às interações de Van der Waals e Coulombicas, e observou-se que, se por um lado, ΔU associado às interações de Van der Waals aumentava ao longo da série, por outro, ΔU associado às interações Coulombicas permanecia constante. Esta observação ainda que feita na transição do estado líquido para o estado gasoso, deixa em aberto que a constância das interações Coulombicas com o aumento da cadeia alquílica possa ser uma realidade num estado mais condensado da matéria como é o caso do estado sólido. Em alguns casos é possível observar que o ponto de fusão oscila quando se vai incrementando a cadeia alquílica substituindo segundo tenha número par ou ímpar de átomos, geralmente apresentando pontos de fusão mais elevados no caso de cadeias ímpares e mais baixos no caso de cadeias pares, esse efeito foi observado em sais de imidazólio (a partir de cadeias acima de 12 carbonos)⁷⁰ ou no caso de sais de triazólio.⁷¹ Nestes casos um número ímpar de átomos de carbono implicaria conformação totalmente *trans* da cadeia permitindo um melhor acomodamento das estruturas na rede cristalina reduzindo a entropia do sistema, no caso de número par de carbonos nem sempre a conformação é totalmente *trans* o que reduz a capacidade de acomodação das estruturas.

No caso do anião algumas observações têm sido feitas relativamente à relação da sua estrutura com o ponto de fusão.⁷³ Normalmente o anião permite pontos de fusão baixos quando há uma maior deslocalização de carga, reduzida possibilidade de estabelecer pontes de hidrogénio com o catião, grande diferença de volumes entre catião e anião (neste caso aumento de entropia), ainda que este último caso não se possa generalizar a todas as situações, pois em átomos mais volumosos a densidade de carga é menor e as consequentes interações iónicas são mais fracas resultando em

redução da entalpia de interacção iónica. O volume do anião não influencia sempre da mesma maneira o ponto de fusão visto haver outras interacções em jogo para além da interacção iónica, já para não falar na entropia de fusão. Para ilustrar essa divergência do efeito do volume no ponto de fusão é de assinalar que no caso de sais de imidazólio o ponto de fusão diminui com o aumento do raio iónico⁷³, enquanto que nos sais de guanidínio o efeito é o reverso.⁷⁴

Todas estas observações são úteis quando os químicos orgânicos pretendem sintetizar um líquido iónico de baixo ponto de fusão estimado qualitativamente por comparação com outros casos semelhantes, mas não tem em conta as particularidades de cada tipo de líquido iónico nem o efeito que pequenas diferenças estruturais podem ter no ponto de fusão (**Figura II.9**).⁷⁵

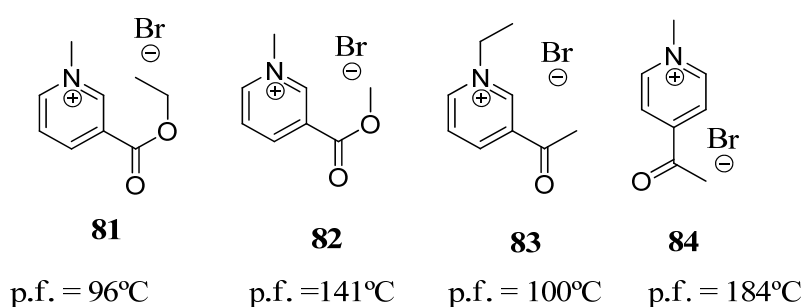


Figura II.9: O efeito que pequenas subtilezas estruturais podem ter no ponto de fusão de sais de piridínio.

Dado o interesse crescente dos líquidos iónicos, rapidamente apareceram métodos/modelos que procuravam prever/explicar de uma forma precisa o seu ponto de fusão recorrendo a parâmetros físico-químicos e/ou parâmetros estruturais. Os primeiros modelos visavam estudar/prever o ponto de fusão com base na variação estrutural dentro de uma determinada família catiónica, mantendo constante o anião usado. Nesse contexto surgiu o estudo de Katritzky *et al*⁷⁶ relativo ao ponto de fusão de 126 brometos de piridínios em que usou a técnica de regressão multilinear (MLR) que resultou numa correlação (R^2) de 0,7683 e RMS de 23,41°C para conjunto de treino de 126 objectos, o modelo foi validado através de validação cruzada (*c.v.*) para três subconjuntos do conjunto de treino com valores de RMS de 27,10, 25,53 e 24,05 °C para cada um dos subconjuntos A, B e C respectivamente. Este modelo foi construído com 6 variáveis relacionadas com a conectividade e ramificação, interacções electrostáticas e graus de liberdade conformacionais e rotacionais do catião. Usando o mesmo conjunto de 126 brometos de piridínio Carrera e Aires de Sousa⁷⁵ usaram conjuntos de árvores de decisão para obter um modelo que previa com maior exactidão os pontos de fusão destes compostos, correlação (R^2) de 0,9598 e RMS de 9,91°C para conjunto de treino, por sua vez em validação cruzada seguindo o mesmo procedimento e os mesmos 3 subconjuntos do conjunto de treino (A, B e C) obtiveram-se valores de RMS de 18,17; 18,70 e 21,96 °C respectivamente, os

descritores mais importantes seleccionados por este método, usando um conjunto inicial de 1085 descritores, estão relacionados com o tamanho da molécula, nomeadamente a presença de cadeias longas, forma e simetria da molécula e presença de sistemas π . Mais recentemente Bini *et al*⁷⁷ utilizaram redes neuronais recursivas para modelar o mesmo conjunto de 126 brometos de piridínio utilizado previamente.^{75,76} O método das redes neuronais recursivas baseia-se na codificação das estruturas de compostos na forma de árvores hierarquizadas, em que cada átomo é um nó, e que posteriormente são utilizadas para fazer regressão. A *performance* deste modelo é inferior à dos anteriores modelos descritos, muito embora o método apresente como vantagem o facto de não ser necessária selecção prévia de descritores. Ao nível de resultados concretos foram obtidos valores de RMS de 30,55, 28,87 e 30,49 °C para conjuntos A, B e C respectivamente como conjuntos de teste em validação cruzada. Para além de brometos de piridínios foram construídos modelos de previsão de pontos de fusão para brometos de imidazólios e brometos de amónios.

No caso de brometos de imidazólios, Katritzky *et al*¹ construíram 4 modelos usando o método de regressão multilinear (MLR), 3 dos modelos referentes a brometos de imidazólio, em que no primeiro conjunto (A – de 57 objectos) as cadeias substituintes nas posições 1 ou 3 do anel de imidazólio eram grupos alquilo, no segundo conjunto (B – de 29 objectos) as posições 1 ou 3 eram grupos OR (R pode ser grupo alquilo, arilo ou H), no terceiro conjunto (C – de 18 objectos) as posições 1 ou 3 eram grupos arilo. Relativamente ao 4º modelo (D – de 45 objectos) compreende brometos de benzoimidazólios (**Figura II.10**):

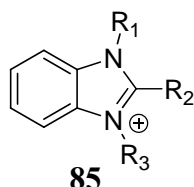


Figura II.10: Estrutura genérica de um benzoimidazólio.

Para o conjunto A de treino o valor de correlação obtido (R^2) foi de 0,7442 e o valor de RMS 29,2°C, o método usado para o conjunto A foi validado usando validação cruzada para 3 subconjuntos (de igual número de elementos), nesse caso a correlação (R^2) para todos os objectos dos 3 subconjuntos foi de 0,7080. Para os conjuntos B, C e D os valores de R^2 oscilaram entre 0,6899 e 0,9432 e os valores de RMS oscilaram entre 14,5 e 20,0°C. Os principais descritores seleccionados por estes modelos basearam-se principalmente nas capacidades de interacções electrostáticas, dimensões e conectividade do catião. Relativamente a brometos de amónios Eike *et al*⁷⁸ construiu 2 modelos, um (A), baseado em 75 brometos de tetraalquil-amónio, e outro (B) baseado em 34 brometos de hidroxialquil-trialquil-amónio. Para construir estes dois modelos foi usado o método de regressão multilinear o valor de R^2 para o conjunto A de treino foi de 0,775 enquanto para conjunto

de B de treino o valor de R^2 foi de 0,716. Descritores que representam a ramificação, flexibilidade, simetria do catião foram seleccionados para o conjunto A. Factores como a presença de ramificações, tamanho das cadeias alquílicas e simetria foram seleccionadas para o conjunto B.

Numa outra abordagem Varnek *et al*⁷⁹ usa conjuntos de 126 brometos de piridínios, 384 brometos de imidazólios e benzoimidazólios, 207 brometos de amónios, construindo modelos separados para cada conjunto e modelos globais, incluindo brometos de imidazólios, amónios e piridínios. Para tal usou diversos métodos entre os quais, Máquinas de suporte vectorial (*Support Vector Machines*, SVM), Regressão Multilinear (multilinear regression, MLR), redes neuronais de retropropagação (*Back Propagation-Neural Networks*, BPG-NN), mínimos quadrados parciais (partial least squares, PLS), entre outros métodos. Foram usados diversos tipos de descritores, entre os quais descritores Dragon,⁸⁰ descritores de fragmentos moleculares e descritores electrotopológicos. Baseados nos resultados obtidos relativos a conjuntos de teste, para a totalidade dos sais (RMS entre 37,5 e 46,4°C), só para brometos de piridínios (RMS entre 26,2 e 34,8°C), só para brometos de imidazólios (RMS entre 38,8 e 45,9°C), e finalmente para brometos de amónios (RMS entre 34,2 e 49,3°C) o autor concluiu que só é possível uma moderada precisão nas previsões de pontos de fusão de líquidos iónicos, estando esse facto relacionado tanto com a qualidade dos dados experimentais como com a complexidade inerente à passagem para o estado líquido mencionada anteriormente.

Para além de líquidos iónicos contendo anião brometo, também foram construídos modelos para outros tipos de aniões. Nesse contexto Sun *et al*⁸¹ obtiveram dois modelos MLR para previsão de pontos de fusão, um para imidazólios contendo tetrafluoroborato como anião (BF_4^-), o outro modelo para imidazólios contendo hexafluorofosfato (PF_6^-) como anião. No primeiro caso usaram-se 16 sais para construir o modelo baseado em três descritores catiónicos (relacionados com a capacidade de formar pontes de hidrogénio, graus de liberdade conformacionais constante dielétrica e momento dipolar do catião), o modelo apresenta um R^2 de 0,9047 e RMS de 19,18 °C para conjunto de treino e R^2 na validação cruzada de 0,7763. No segundo modelo foram usados 25 sais baseados no anião hexafluorofosfato para construir modelo baseado em seis descritores moleculares (relacionados com a capacidade de formar pontes de hidrogénio, carga mínima associada ao átomo de hidrogénio e interacções intramoleculares envolvendo átomos de azoto no anel de imidazólio), o modelo apresenta um R^2 de 0,9207 e RMS de 15,23 °C para conjunto de treino e R^2 na validação cruzada de 0,8423. Até agora os modelos de previsão de pontos de fusão de líquidos iónicos baseavam-se na variação da estrutura do catião, permanecendo o anião constante.

López-Martin *et al*⁸² estudaram o efeito do anião e do catião no ponto de fusão de líquidos iónicos e para tal usaram o método PLS. No primeiro estudo mantiveram constante o catião (1-etil-3-metilimidazólio) e foram variando a estrutura do anião; foram usados 22 sais nessas condições. Os descritores mais relevantes para descrever o ponto de fusão estão relacionados com o tamanho,

distribuição de carga e simetria do anião (tal como no caso do catião em estudos anteriores). No segundo estudo variaram tanto o catião de imidazólio como o anião, foram usados 62 líquidos iónicos para construir o modelo do qual resultou um R^2 de 0,869. Os descritores mais relevantes para este modelo são fundamentalmente catiónicos relacionados com o tamanho, simetria e carga; relativamente ao anião o grau de esfericidade foi usado pelo modelo como um descritor relevante. Estes dois modelos não foram validados interna ou externamente.

Numa abordagem completamente distinta Krossing *et al*⁸³ usam equações termodinâmicas usando parâmetros determinados por cálculos quânticos e valores experimentais de constante dielétrica (ϵ) para determinar a energia de Gibbs (kJ.mol^{-1}) a condições de pressão e temperatura padrão (25°C, 10^5 Pa) para a fusão (ΔG_f°). E assim prever e explicar o facto de um líquido iónico ser líquido à temperatura ambiente, tendo em conta que o valor de ΔG_f à temperatura de fusão é 0 e que nas condições padrão ΔG_f° tem valor negativo para líquidos à temperatura ambiente e valor positivo para sólidos à temperatura ambiente. Para determinar valor de ΔG_f° o autor recorre ao ciclo de Born-Fajans-Haber (**Figura II.11**):

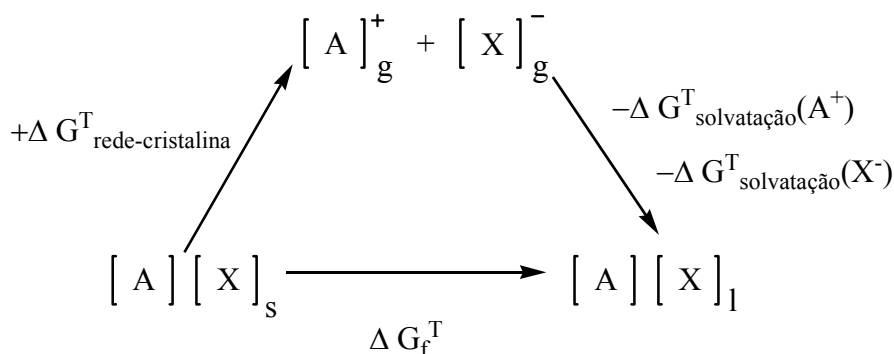


Figura II.11: Ciclo de Born-Fajans-Haber para determinar ΔG_f^T .

Deste ciclo resulta que:

$$\Delta G_f^T = \Delta G_{\text{rede-cristalina}}^T - \Delta G_{\text{solvatação}}^T(A^+) - \Delta G_{\text{solvatação}}^T(X^-)$$

Equação II.17

em que a energia de Gibbs de fusão é resultado da diferença entre a energia de Gibbs de rede cristalina e as energias de Gibbs de solvatação para catião e anião. Tudo isto para condições genéricas de temperatura T . Daqui em diante, por comodidade, serão consideradas condições de pressão e temperatura padrão. Decompondo a **Equação II.17**:

$$\Delta G_{rede-cristalina}^{\circ} = \Delta H_{rede-cristalina}^{\circ} - T \cdot \Delta S_{rede-cristalina}^{\circ}$$

Equação II.18

em que $\Delta G_{rede-cristalina}^{\circ}$ é resultado da diferença entre a entalpia de rede cristalina e o produto da temperatura pela entropia de rede cristalina. O parâmetro $\Delta H_{rede-cristalina}^{\circ}$ da **Equação II.18** é obtido a partir seguinte expressão a pressão constante:

$$\Delta H_{rede-cristalina} = U + 2RT$$

Equação II.19

em que R é a constante dos gases perfeitos ($\text{kJ.K}^{-1}.\text{mol}^{-1}$). Para determinar $\Delta H_{rede-cristalina}$ (para substituir na **Equação II.18**) é necessário saber a energia potencial U (kJ.mol^{-1}) que se calcula empiricamente através da seguinte equação válida para uma série de sais inorgânicos do tipo AX(1:1) obtida por Jenkins *et al*:⁸⁴

$$U = 2 \cdot \left(\frac{\alpha}{\sqrt[3]{V_m}} + \beta \right)$$

Equação II.20

em que α e β são coeficientes empíricos (117,3 e 51,9 kJ mol^{-1} respectivamente) e V_m (nm^3) o volume da formula molecular do sal (líquido iônico).

Para calcular o valor de V_m do sal ora se recorre aos dados cristalográficos baseados no volume da célula unitária, ou somando os volumes atômicos médios dos elementos presentes na fórmula molecular do sal. O termo entrópico $\Delta S_{rede-cristalina}^{\circ}$ ($\text{kJ.K}^{-1}.\text{mol}^{-1}$) da **Equação II.18** determina-se pela diferença entre a entropia do sal no estado gasoso e a entropia do sal na rede cristalina:

$$\Delta S_{rede-cristalina} = \sum S^{\circ}(A^{+} + X^{-})_g - S^{\circ} AX_s$$

Equação II.21

em que o parâmetro $\sum S^{\circ}(A^{+} + X^{-})_g$ se determina recorrendo a cálculos quânticos para o estado gasoso. Relativamente ao termo entrópico associado ao estado sólido, este pode ser calculado recorrendo à expressão empírica geral determinada por Jenkins *et al*:⁸⁵

$$S^{\circ} = kV_m + c$$

Equação II.22

em que V_m é o volume molecular e k e c constantes empíricas ($1360 \text{ J K}^{-1} \text{ nm}^{-3}$ e $15 \text{ J K}^{-1} \text{ mol}^{-1}$ respectivamente) válidas para sais inorgânicos. Assim, determina-se o valor de $\Delta G_{\text{rede-cristalina}}^o$. Para saber o valor de ΔG_f^o é necessário saber ainda o valor de $\Delta G_{\text{solvação}}^o$ para catião e anião, que podem ser determinados através de cálculos quânticos, para esse cálculo é necessário o valor da constante dielétrica ϵ , que no caso deste estudo foi obtida experimentalmente através de espectroscopia dielétrica. Relativamente aos 14 líquidos iónicos estudados, todos tinham valor de ΔG_f^o negativo e experimentalmente eram todos líquidos à temperatura ambiente, logo o modelo foi capaz reconhecer líquidos iónicos à temperatura ambiente. Com este modelo podia ser possível prever o valor de ponto de fusão prevendo ΔG_f^T para uma série de temperaturas e fazendo um gráfico ΔG_f^T vs T descobrir a temperatura à qual ΔG_f iguala 0, essa temperatura é o ponto de fusão. Mas para tal teria que se partir de dois pressupostos, o primeiro que ϵ seja independente da temperatura (na prática as variações são pouco significativas com a temperatura) e que $\Delta G_{\text{solvação}}^T$ seja independente da temperatura, o que não acontece visto estar dependente de um termo entrópico dependente da temperatura. Logo para determinar o valor do ponto de fusão terá que se partir de novo ciclo de Born-Fajans-Haber, agora modificado (**Figura II.12**):

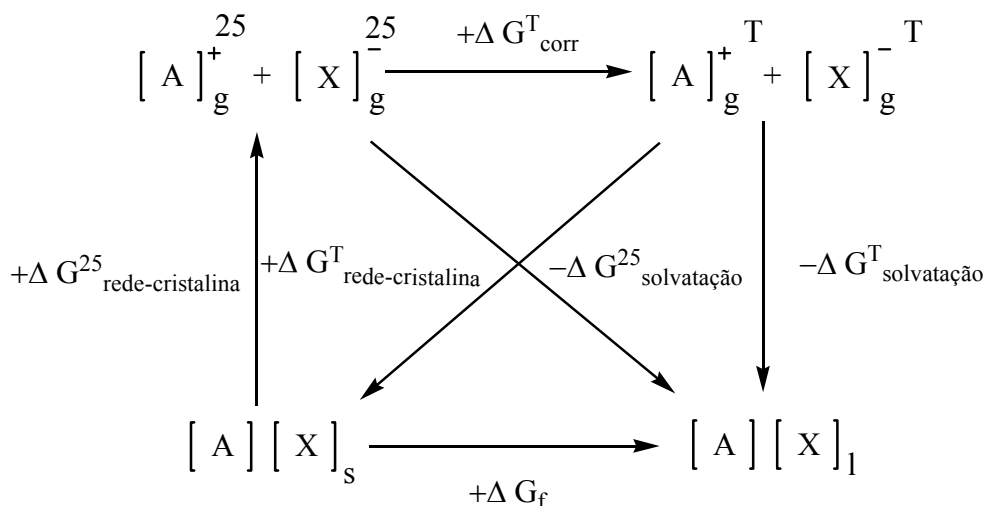


Figura II.12: Ciclo Born-Fajans modificado para cálculo de ΔG_f^T .

Deste ciclo modificado resulta que a uma determinada temperatura T diferente da temperatura padrão (25°C) a energia de Gibbs de fusão (kJ.mol^{-1}) é definida por:

$$\Delta G_f^T = \Delta G_{\text{rede-cristalina}}^T - \Delta G_{\text{solvação}}^T$$

Equação II.23

Como estes cálculos de mecânica quântica para $\Delta G_{\text{solvação}}^T$ são feitos nas condições padrão de temperatura, não é possível determinar este valor directamente, mas usando este ciclo de Born-Fajans-Haber modificado é possível esta determinação indirecta:

$$\Delta G_{\text{solvatação}}^T = \Delta G_{\text{solvatação}}^{25} + \Delta G_{\text{corr}}^T$$

Equação II.24

Substituindo a **Equação II.24** na **Equação II.23** obtêm-se:

$$\Delta G_f^T = \Delta G_{\text{rede-cristalina}}^T - \Delta G_{\text{solvatação}}^{25} + \Delta G_{\text{corr}}^T$$

Equação II.25

Em que ΔG_{corr}^T se pode determinar da seguinte maneira:

$$\Delta G_{\text{corr}}^T = \Delta G_{\text{rede-cristalina}}^T - \Delta G_{\text{rede-cristalina}}^{25}$$

Equação II.26

No cálculo de $\Delta G_{\text{rede-cristalina}}^T$ usa-se como aproximação a entropia padrão do sal (a 25°C) usada nas **Equações II.21** e **II.22**, de qualquer das formas espera-se que este erro seja reduzido. Desta forma foi possível estimar o ponto de fusão para 14 líquidos iônicos (presentes 4 tipos de aniões diferentes e 9 catiões diferentes baseados nas unidades de imidazólio, amónio, pirrolidínio e piridínio) com os valores previstos a aproximarem-se dos valores experimentais.

Este constitui um método diferente do habitual e bastante elegante que possui como desvantagens a necessidade de obter experimentalmente os valores das constantes dielétricas e o facto de contabilizar apenas as interações associadas aos sais inorgânicos. O método foi testado em apenas 14 líquidos iônicos.

No trabalho desta tese serão usados métodos automáticos de determinação de ponto de fusão de líquidos iônicos, recorrendo exclusivamente a características estruturais de cada sal para estabelecer a relação com o ponto de fusão, dentro de um tipo de catiões será contabilizado o efeito de grupos substituintes nesse catião tal como o efeito de diferentes aniões.

II - 1.3.2: Densidade

A densidade é uma propriedade fundamental de todos os materiais e está definida pela massa que uma substância possui por unidade de volume:

$$\rho = \frac{m}{v}$$

Equação II.27

cujas unidades SI são o kg.m^{-3} , ainda que seja corrente o uso das unidades g.cm^{-3} . A densidade é uma medida da quantidade de matéria por unidade de volume, reflectindo também as interacções e o distanciamento entre as diversas estruturas químicas. Esta propriedade define a aplicabilidade de uma substância em diversas situações reais, tais como permitir que um iceberg ou um navio flutuem, em química, definir num sistema extractivo que solvente irá ficar na fase superior ou qual irá ficar na fase inferior, com implicações tanto ao nível de síntese laboratorial como ao nível industrial relativamente à definição do sistema e à concepção do equipamento; permite também definir a eficácia de um material energético,⁹⁶ como é o caso dos explosivos, estando dependentes da densidade ρ , a pressão de detonação P e a velocidade de detonação D . Ainda no caso de materiais energéticos, mais concretamente sais energéticos, a densidade contribui para definir a entalpia de formação $\Delta H_{\text{formação}}^{\circ}$ (kJ.mol^{-1}) de um composto resultado da soma das entalpias de formação de catião $\Delta H_{\text{formação-catião}}^{\circ}$, anião $\Delta H_{\text{formação-anião}}^{\circ}$ e entalpia de rede cristalina $\Delta H_{\text{rede-cristalina}}$:

$$\Delta H_{\text{formação}}^{\circ} = \Delta H_{\text{formação-catião}}^{\circ} + \Delta H_{\text{formação-anião}}^{\circ} + \Delta H_{\text{rede-cristalina}}$$

Equação II.28

Em que $\Delta H_{\text{rede-cristalina}}$, (anteriormente usada em II - 1.3.1 para determinar o ponto de fusão de líquidos iónicos) é função da densidade ρ (g.cm^{-3}) do composto:

$$\Delta H_{\text{rede-cristalina}} = 1981,2 \left(\sqrt[3]{\frac{M}{\rho}} \right) + 108,8$$

Equação II.29

em que M é a massa molar do composto (g.mol^{-1}). Ainda no campo dos parâmetros físico-químicos, a densidade ρ permite a determinação da tensão superficial σ ⁸⁶ (mJ.m^{-2}) através da seguinte equação empírica:

$$\sigma = (C.M.\rho)^4$$

Equação II.30

em que, C é uma constante empírica e M a massa molar do composto. Existe também uma relação entre densidade e índice de refração R_1 definida por:⁸⁶

$$R_M = \left(\frac{C.M}{\rho} \right) \left(\frac{R_I^2 - 1}{R_I^2 + 2} \right)$$

Equação II.31

em que C é uma constante empírica, M a massa molar do composto e R_M a refractividade molar. É ainda possível determinar o coeficiente de expansão térmica α_p (K^{-1}) e a compressibilidade isotérmica k_T (Pa^{-1}), a partir da densidade:

$$\alpha_p = -\frac{1}{\rho} \left(\frac{\partial \rho}{\partial T} \right)_p$$

Equação II.32

$$k_T = \frac{1}{\rho} \left(\frac{\partial \rho}{\partial P} \right)_T$$

Equação II.33

em que T (K) é a temperatura e P a pressão (Pa).

Dada a importância da densidade cedo se procurou relacionar a estrutura de um composto com a sua densidade. Ao nível dos alcanos observou-se que a densidade aumenta à medida que aumenta o número de átomos de carbono na cadeia, por sua vez os alcanos cíclicos possuem densidades mais elevadas que os respectivos alcanos de cadeia linear.⁸⁷ No caso de moléculas mais complexas torna-se complicado fazer uma associação directa entre estrutura e densidade exacta surgindo modelos/metodologias mais ou menos gerais, mais ou menos automáticas. Toporov e Toporova,⁸⁸ partindo de um conjunto variado de alcanos, usaram descritor que combina informação das orbitais atómicas de cada elemento (GAO) e informação relativa à conectividade de cada elemento, construindo uma simples regressão linear obtendo um coeficiente de correlação R de 0,9922 para conjunto de treino de 66 moléculas e R de 0,988 para conjunto de teste de 67 moléculas. Numa abordagem diferente, numa aplicação mais generalizada Hofmann⁸⁹ determina os volumes atómicos de 89 elementos baseados nos dados cristalográficos de volume de um cristal para uma série de 182 239 compostos, resolvendo um sistema de equações baseado na seguinte equação para o volume V do cristal de um determinado composto:

$$V = \sum_{i=1}^{100} n_i v_i (1 + \alpha T)$$

Equação II.34

Em que n_i é o número de ocorrências de um determinado elemento i , v_i o volume médio desse elemento, α o coeficiente de expansão térmico médio e T a temperatura, em que 100 no somatório é o número de elementos diferentes considerados no início do estudo. O cálculo do volume médio tem um erro de 0,3%. Este método de determinação de volume sobreavalia o volume ocupado por compostos iónicos. Ainda assim não deixa de ser um método geral para determinar o volume ocupado por um composto e consequentemente poder calcular o valor de densidade, visto ser incomparavelmente mais fácil aceder à massa de um composto. Jenkins *et al*⁸⁴ calculam volumes para sais através da expressão:

$$V = p.V_{cat} + q.V_{an}$$

Equação II.35

em que V é o volume total do sal, V_{cat} e V_{an} são o volume de catião e anião respectivamente, p e q são as proporções com que os componentes de anião e catião se combinam respectivamente para formar o sal. Com esta relação muito se pode fazer, sabendo os volumes de simples catiões ou de aniões como os de metais alcalinos ou halogéneos, é possível combinar os dados de volume de sal, derivado de dados cristalográficos

$$V = a.b.c. \frac{\sqrt{(1 - \cos^2 \alpha - \cos^2 \beta - \cos^2 \gamma + 2.\cos \alpha.\cos \beta.\cos \gamma)}}{Z}$$

Equação II.36

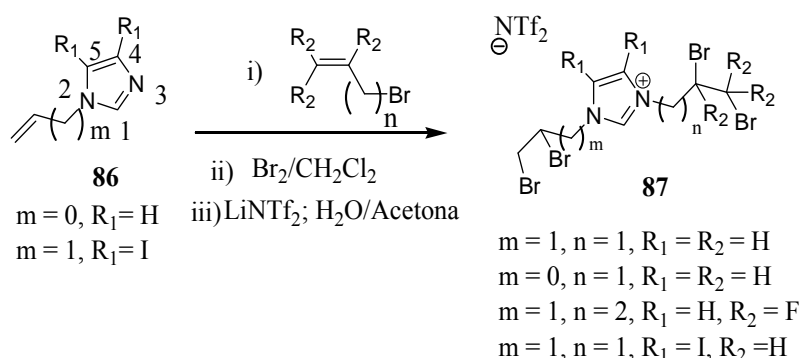
baseados nas dimensões a , b , c , ângulos α , β , γ e número de moléculas Z na célula unitária, para um número relativamente reduzido de sais e derivar volumes para catiões e aniões mais complexos, podendo obter exclusivamente a partir de tabelas, construídas com volumes para catião e anião, determinar o volume de um novo sal.

II - 1.3.2.1: Densidade de Líquidos Iónicos: Perspectiva Sintética

Relativamente aos líquidos iónicos, dadas as suas inúmeras aplicações, nomeadamente como meio reaccional, surgiu o interesse em observar empiricamente a variação da densidade em função da estrutura de catião e anião. Uma observação bastante relevante, efectuada por Esperança *et al*⁹⁰ na família de 1-aquil-3-metilimidazólios consistia no facto que por cada dois grupos $-\text{CH}_2-$ que se inseria na cadeia alquílica o volume aumentava invariavelmente $34,4 \pm 0,5 \text{ cm}^3.\text{mol}^{-1}$ independentemente do anião usado, esta linearidade e independência demonstradas facilitam a determinação de volumes de outros aniões e catiões. Este aumento de volume verificado ao longo da série dos 1-alquil-3-metilimidazólios resulta num decréscimo progressivo da densidade. Noutro estudo observou-se que a presença de flúor na estrutura do líquido iónico, tanto no anião como no catião provocam um aumento

da densidade.⁹¹ Outros halogénios provocam aumento pronunciado da densidade, tal como o iodo e o bromo,⁹² visto os halogénios serem elementos substancialmente mais pesados que átomos de hidrogénio. Ainda falando sobre halogénios, a presença de grupos pentafluorosulfanos (-SF₅) no catião provoca em geral um aumento de densidade, ao mesmo tempo que mantém o ponto de fusão num valor reduzido.⁹³ Também se observa em geral que a presença de núcleo catiónico aromático conduz a densidades mais elevadas que núcleos alifáticos, em grande medida devido às interações π - π que reduzem as distâncias intermoleculares. Devido ao propiciar interações por pontes de hidrogénio, os grupos nitrilo no catião, têm potencial para aumentar a densidade⁹⁴ principalmente quando o anião tem a mesma capacidade para formar essas interações. Ter estrutura cíclica e ser simétrico são características que o catião deve ter para que aumente a densidade, pois reduzidos graus de liberdade conformacionais e rotacionais contribuem para a redução das distâncias intermoleculares, reduzindo-se o volume livre entre estruturas. Relativamente ao anião, este influencia muito significativamente a densidade, da mesma forma que a presença de flúor no catião aumentava a densidade, também no anião isso acontece, sendo, linhas gerais $[\text{CH}_3\text{SO}_3]^- \approx [\text{BF}_4]^- < [\text{CF}_3\text{CO}_2]^- < [\text{CF}_3\text{SO}_3]^- < [\text{PF}_6]^- < [(\text{CF}_3\text{SO}_2)_2\text{N}]^-$ a ordem de densidades para líquidos iônicos contendo estes aniões. A presença de elementos pesados no anião contribui para um acréscimo muito significativo da densidade como é o caso de sais contendo o anião oxipentafluorotungstato ($[\text{WOF}_5]^-$).⁹⁵

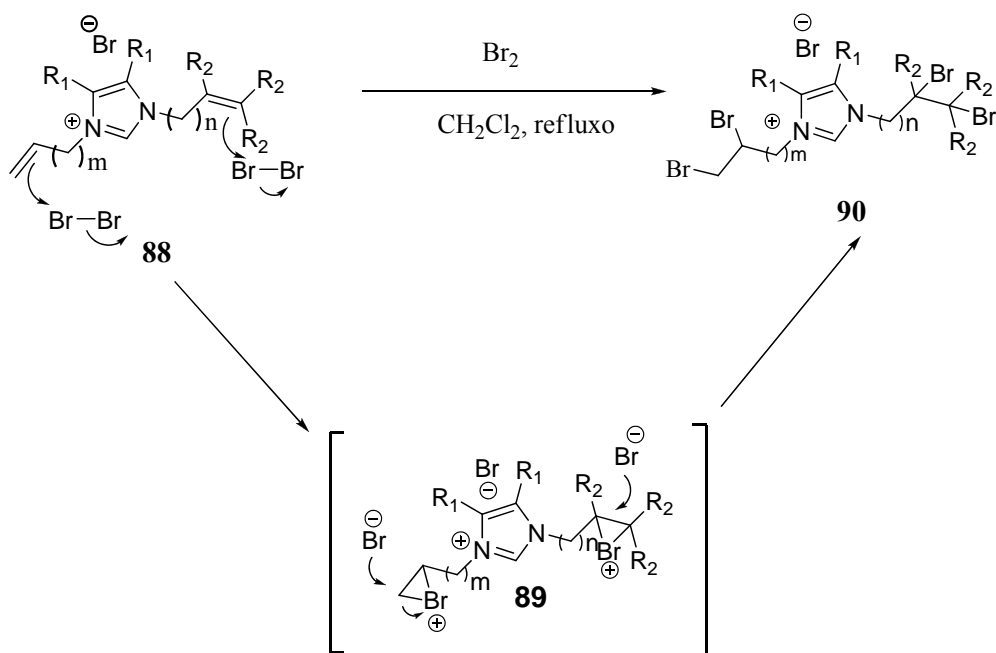
O interesse em líquidos iônicos mais densos surge, não só no contexto dos já mencionados materiais energéticos ou como sistema reaccional ou extractivo, mas também no contexto de fluidos inerciais usados em instrumentos de navegação nomeadamente acelerómetros e giroscópios. Estes fluidos são normalmente voláteis e tóxicos e o uso de líquidos iônicos pode mitigar esse problema, nesse contexto surgiu o trabalho de síntese de Ye e Shreeve⁹² que, partindo de imidazóis, obtiveram diversos sais de imidazólio bromados. Surge aqui um esquema típico do método sintético utilizado:



Esquema II.15: Método de síntese para obtenção de sais de imidazólio bromados.

Neste esquema de síntese distinguem-se três passos distintos, o primeiro passo consiste numa substituição nucleofílica S_N2 quando n do reagente é 2 e S_N2 ou S_N1, quando n no reagente é 1,

dependendo da nucleofilicidade do imidazol, resultado do ataque nucleofílico do azoto 3 do imidazol no carbono α ao bromo no reagente *i*. O segundo passo da síntese, consiste na bromação das posições olefinicas do sal de imidazólio:



Esquema II.16: Bromação olefinica de sal de imidazólio.

Neste caso representou-se o ataque em simultâneo das duas olefinas com a formação de um único intermediário tendo em vista a simplificação do esquema apresentado (**Esquema II.16**). Assim, como resultado do ataque nucleofílico de olefina ao bromo forma-se o intermediário bromônio que sofre ataque *anti* de anião brometo resultando no sal de imidazólio bromado. No terceiro e último passo irá ocorrer troca de anião adicionando bis(trifluorometilsulfonil)imidato de lítio à mistura reaccional, ocorrendo a troca do anião brometo pelo anião bis(trifluorometilsufonil)imidato (NTf_2^-) resultando em líquidos iônicos com densidades compreendidas entre 1,95 e 2,80 g/cm^3 . Inspirado por esse trabalho,⁹² foram sintetizados e estudados ao longo desta dissertação novos sais bromados, mas agora são sais de guanidínio.

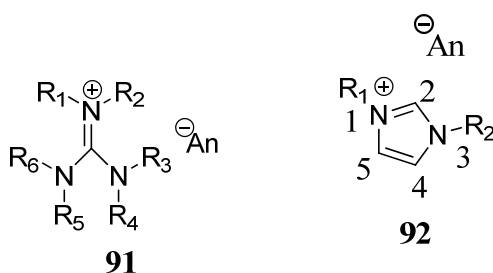


Figura II.13: Sal de guanidínio vs sal de imidazólio.

Os sais de guanidínio têm mais cadeias substituintes que os habitualmente usados sais de imidazólio 1,3-disubstituídos (**Figura II.13**), logo aqueles têm maiores potencialidades para possuir mais posições olefinicas que podem ser bromadas, podendo assim produzir um sal com maior número de bromos. Assim foi estudada forma de sintetizar sais de guanidínio bromados, e investigada a forma como a densidade é influenciada pela presença de halogênios, tanto no catião como no anião.

II - 1.3.2.2: Densidade de Líquidos Iônicos: Métodos de Previsão

Empiricamente é possível relacionar a densidade com a estrutura de um líquido iônico e saber qualitativamente se um líquido iônico vai ser denso ou não, mas ainda assim o valor quantitativo da densidade permanece inacessível, nesse contexto surgiram métodos/modelos específicos de previsão de densidade de líquidos iônicos, com maior ou menor grau de sofisticação. Ye e Shreeve⁹⁶ usando uma metodologia semelhante à utilizada por Jenkins *et al*⁸⁴ para sais inorgânicos, determinando os volumes de catião e anião para líquidos iônicos a partir de algumas tabelas de volumes para iões simples, alguns dados cristalográficos e a relação simplificada entre o volume do sal e o volume de catião e anião expressa na **Equação II.35**. Adicionalmente determinam o volume de um ião não tabelado pela soma ou diferença de funcionalidades específicas (estas já tabeladas) dentro do ião, usando os dados de volume determinados para prever a densidade que um determinado líquido iônico vai ter. Ao nível de resultados o método permite obter um erro médio absoluto (MAE) de 0,020 g.cm⁻³ para um conjunto de 35 imidazólios líquidos à temperatura ambiente, e um MAE de 0,035 g.cm⁻³ para conjunto de 150 sais energéticos. Usando método completamente distinto Deetlefs *et al*⁸⁶ usam num método a tensão superficial para determinar a densidade, tal como o índice de refração para o mesmo fim. No primeiro método partem de uma simplificação da equação proposta por Macleod⁹⁷ que em 1923 relacionou a densidade com a tensão superficial:

$$C.\rho = \sigma^{\frac{1}{4}}$$

Equação II.37

em que ρ é a densidade, σ a tensão superficial e C uma constante. Modificando esta expressão multiplicando cada lado da equação pela massa molar do composto M e rearranjando-a irá resultar:

$$C.M = \frac{M.\sigma^{\frac{1}{4}}}{\rho}$$

Equação II.38

O produto $C.M$ é a denominada tensão superficial pesada segundo o volume molar (*Parachor*). Este produto, que estabelece a relação entre densidade e tensão superficial, é aditivo

relativamente a componentes estruturais, o que facilita a sua determinação, assim é necessário apenas o valor da tensão superficial para obter a densidade. Outra forma de determinar a densidade ρ é através do índice de refração R_I , refratividade molar R_M e a massa molar do composto M através da relação de Lorentz-Lorenz:

$$R_M = \frac{M}{\rho} \left(\frac{R_I^2 - 1}{R_I^2 + 2} \right)$$

Equação II.39

A refratividade molar R_M , à semelhança do *Parachor*, é aditiva e pode ser determinada através de componentes estruturais tabelados, logo seria apenas indispensável saber o índice de refração para determinar a densidade.

Através das **Equações II.38 e II.39** foi possível obter excelentes correlações para conjunto limitado de sais de imidazólio (9 líquidos iônicos), $R^2 = 0,999$ e $R^2 = 0,998$ respectivamente. Numa outra abordagem Trohalaki *et al*⁷¹ usam métodos heurísticos baseados em diversas regressões lineares para obter a densidade de brometos de triazólio, os descritores usados representam fundamentalmente o grau de interação electrostática e a relação entre a carga superficial associada a interações por pontes de hidrogénio e a superfície total. Uma metodologia mais sofisticada foi usada por Palomar *et al*⁹⁸ para determinar a densidade de 40 sais de imidazólio em que também varia o anião utilizado, neste caso usando método de química quântica (COSMO) para, primeiro, otimizar a geometria do sal e, segundo, determinar as diferentes interações entre catião e anião, função das distribuições de polaridade na superfície molecular, de forma a determinar o volume do sal e assim determinar a densidade. Para os 40 sais de imidazólio conseguiram um R^2 de 0,995 comparando os valores previstos com os valores experimentais. Este método permite obter a distribuição de carga sigma para um dado sal, sendo possível distinguir uma zona de interações polares e outra de interações apolares, e através dessa distribuição fazer uma interpretação das densidades observadas, no caso de imidazólios de cadeia curta observou-se uma distribuição mais simétrica de cargas relativamente a catião e anião. À medida que a cadeia ia aumentando a distribuição de cargas era mais assimétrica podendo resultar em maiores interações repulsivas entre zonas polares e apolares do sal. Neste trabalho são usados diversos métodos automáticos para previsão de densidade de sais de imidazólio em que para além da variação estrutural no catião é tida em conta a variação estrutural no anião. Dadas as observações aqui mencionadas de grande independência volumétrica entre catião e anião, estes dois são tratados separadamente neste trabalho.

II-2: Metodologia Computacional

No trabalho descrito neste capítulo recorreu-se a quatro conjuntos de dados distintos que foram utilizados na construção de quatro tipos diferentes de modelos: dois modelos referentes à previsão de ponto de fusão/estado físico de sais de imidazólio, um modelo de previsão de pontos de fusão de sais de guanidínio e um modelo de previsão de densidade de sais de imidazólio. Na construção desses modelos foram usados, como métodos de aprendizagem automática, redes neurais de *Counter-Propagation*, Árvores de Decisão, *Random Forest* e *Support Vector Machines*. Estes modelos para além de serem validados interna e externamente foram ainda utilizados para orientar a síntese de novos sais, potenciais líquidos iônicos.

II-2.1: Conjuntos de Dados

Os dados usados nestes estudos foram recolhidos directamente da literatura científica, concretamente foram recolhidas estruturas de sais e as suas respectivas propriedades físico-químicas de interesse disponíveis, como ponto de fusão (ou transição vítrea), estado físico à temperatura ambiente e densidade. Os dados recolhidos foram agrupados em quatro conjuntos referentes ao tipo de catião e à propriedade físico-química em questão: **a)** conjunto de sais de imidazólio e seus respectivos pontos de fusão; **b)** conjunto de sais de imidazólio classificados segundo uma de três classes, sendo considerados sais que sejam líquidos a partir de temperaturas até 25°C (classe a), a temperaturas entre 25 e 100°C (classe b) e a temperaturas superiores a 100°C (classe c); **c)** conjunto de sais de imidazólio e suas respectivas densidades; **d)** conjunto de sais de guanidínio e os seus respectivos pontos de fusão.

II-2.1.1: Representação das Estruturas Químicas por Descritores

Moleculares

Considerando o conjunto de dados concreto e o método de aprendizagem automático seleccionado, foram usados diferentes descritores moleculares para representação da estrutura de cada sal ou catião. Para a obtenção dos descritores moleculares a estrutura 3D de cada catião e de cada anião (quando caso disso) foi optimizada usando o programa CORINA 3.0.^{99,100} A estrutura 3D optimizada foi submetida a diferentes programas de cálculo. O programa PETRA 3.1¹⁰¹ foi utilizado para associar propriedades físico-químicas a cada átomo, ligação e à estrutura como um todo. O cálculo dessas propriedades foi utilizado para obtenção de diversos descritores associados à estrutura de um catião ou anião. A partir das propriedades calculadas pelo PETRA 3.1 foram calculados descritores RDF¹⁰² obtidos pelo programa RCODE, descritores de superfície acessível determinados pelo programa SURFACE¹⁰³ e descritores variados determinados pelo programa XPET. Antes de poder calcular descritores RDF e SURFACE houve uma selecção prévia de propriedades físico-químicas a associar a esses tipos de descritores, essa selecção consistiu no uso dum conjunto de 995

descritores contendo conjuntos de 128 descritores RDF, cada conjunto a ser calculado tendo em conta uma determinada propriedade físico-química, o mesmo acontecendo com descritores SURFACE; o conjunto de descritores ficou completo com descritores XPET do catião e descritores que representassem o anião. Posteriormente foi construído um modelo *Random Forest* com todos os 995 descritores que permitiu a obtenção de uma medida de importância associada a cada descritor. Para cada conjunto específico de descritores, calculados segundo uma propriedade físico-química concreta, foi calculada uma média dos valores de importância, no final aquela propriedade físico-química que permitia a obtenção da média de importância mais elevada era a propriedade seleccionada para associar aos descritores RDF, o mesmo acontecendo com os descritores SURFACE. Até agora foram focados descritores derivados do programa PETRA sendo adicionalmente calculados descritores DRAGON,¹⁰⁴ partindo-se para esse cálculo da estrutura tridimensional otimizada pelo programa CORINA 3.0.

a) Conjunto de Sais de Imidazólio e Respectivo Ponto de Fusão

Neste caso usaram-se descritores tanto para representar o catião como o anião. Usaram-se 71 descritores aniônicos compreendendo 45 descritores XPET e 26 descritores binários. Os descritores binários têm como base um conjunto de 6 descritores cada um referente à presença de um determinado anião na estrutura do sal sendo representada a sua presença pelo valor “1” e a sua ausência pelo valor “0”. Outros 20 descritores binários derivam dos 6 descritores originais multiplicando estes pelo peso molecular do catião, peso molecular do anião e diferença entre peso molecular de catião e anião.

Relativamente aos descritores que representam o catião estes são 210 e incluem 128 descritores RDF que dão a distribuição de distâncias inter-atómicas dentro do catião, sendo pesadas pela polarizabilidade atómica associada a cada átomo. Nesta representação não estão incluídos átomos de hidrogénio. Adicionalmente foram usados 37 descritores SURFACE que representam a superfície acessível do catião sendo determinados tendo em conta a electronegatividade sigma associada a cada átomo.

Finalmente foram usados 45 descritores XPET. Aparecem representados em seguida (**Tabela II.1**) todos os 281 descritores utilizados na construção destes modelos. No caso do modelo *Random Forest* foram utilizados os 281 descritores base (todos os descritores representados na **Tabela II.1**), enquanto que no caso do modelo baseado em *Support Vector Machines* foram utilizados subconjuntos dos descritores mais importantes seleccionados por modelo *Random Forest*, sendo o critério de selecção o incremento da percentagem do valor de MSE quando determinado descritor assumia valores aleatórios. O subconjunto de descritores seleccionado consistiu nos 30 descritores mais importantes seleccionados pelo modelo *Random Forest* (descritor 14, 16, 17, 24, 25, 28, 29, 30, 31,

35, 37, 41, 53, 70, 130, 131, 180, 181, 196, 198, 204, 205, 206, 226, 253, 262, 274, 277, 280 e 281 - **Tabela II.1**), sendo aquele subconjunto que proporcionava R^2 mais elevados para conjunto de treino segundo validação cruzada, partindo conjunto de treino em 5 subconjuntos. No caso do modelo baseado em Árvore de Decisão foram seleccionados 11 descritores (descritor 1, 25, 28, 35, 37, 43, 70, 91, 180, 226 e 281 - **Tabela II.1**).

Tabela II.1-1: Descritores base usados no estudo de modelação de pontos de fusão de sais de imidazólio.

Descritor número	Ião	Descritor	Descrição
1	Catão	RAPsH1	Descritor RDF considerando a polarizabilidade de átomos (sem hidrogénios incluídos) que distem entre si perto de 0,1 Å
2	""	RAPsH2	"" 0,2 Å
3-127	""	-	Descritores RDF considerando a polarizabilidade de átomos (sem hidrogénios incluídos) que distem entre si 0,3 – 12,7 Å
128	""	RAPsH128	Descritor RDF considerando a polarizabilidade de átomos (sem hidrogénios incluídos) que distem entre si perto de 12,8 Å
129	""	SENSIG1	Descritor SURFACE considerando pontos a distar entre si 0-1 Å
130	""	SENSIG2	"" 1-2 Å
131-164	""	SENSIG3 - SENSIG36	Descritores SURFACE considerando pontos a distar entre si 2-3 Å a 35-36 Å
165	""	SENSIG37	Descritor SURFACE considerando pontos a distar entre si 36-37 Å
166	""	XP1	ATOMS – Número de átomos
167	""	XP2	BONDS – Número de ligações
168	""	XP3	MW – Massa molar
169	""	XP4	ALIPH IND – Índice alifático
170	""	XP5	AROM IND – Índice aromático
171	""	XP6	H_ACC – Potencial máximo de aceitador de hidrogénio
172	""	XP7	H_DON – Potencial máximo de doador de hidrogénio
173	""	XP8	MMPOL – Polarizabilidade
174	""	XP9	NH – Número de NH
175	""	XP10	NH2 – Número de NH ₂
176	""	XP11	HEAVY – Número de átomos pesados
177	""	XP12	OH – Número de grupos hidroxilo
178	""	XP13	NrO – Número de oxigénios
179	""	XP14	NrN – Número de azotos
180	""	XP15	QMIN – Carga mínima
181	""	XP16	QMAX – Carga máxima
182	""	XP17	MAX_CHAIN – Cadeia mais longa
183	""	XP18	N_CHAINS – Número de cadeias
184	""	XP19	DIST_CH[1]
185	""	XP20	DIST_CH[2]
186	""	XP21	DIST_CH[3]
187	""	XP22	DIST_CH[4]
188	""	XP23	DIST_CH[5]
189	""	XP24	DIST_CH[6]

Tabela II.1-2: Continuação da apresentação dos descritores base usados no estudo de modelação de pontos de fusão de sais de imidazólio.

Descritor número	Ião	Descritor	Descrição
190	Catão	XP25	DIST_CH[7]
191	"	XP26	DIST_CH[8]
192	"	XP27	DIST_CH[9]
193	"	XP28	DIST_CH[10]
194	"	XP29	DIST_CH[11]
195	"	XP30	DIST_CH[12]
196	"	XP31	DIST_CH_TOP[1]
197	"	XP32	DIST_CH_TOP[2]
198	"	XP33	DIST_CH_TOP[3]
199	"	XP34	DIST_CH_TOP[4]
200	"	XP35	DIST_CH_TOP[5]
201	"	XP36	DIST_Q[1]
202	"	XP37	DIST_Q[2]
203	"	XP38	DIST_Q[3]
204	"	XP39	DIST_Q[4]
205	"	XP40	DIST_Q[5]
206	"	XP41	DIST_Q[6]
207	"	XP42	DIST_Q[7]
208	"	XP43	DIST_Q[8]
209	"	XP44	DIST_Q[9]
210	"	XP45	DIST_Q[10]
211	Anião	XP1	ATOMS – Número de átomos
212	"	XP2	BONDS – Número de ligações
213-255	"	XP3-XP45	O mesmo que descritores 168 - 210
256	"	B1	Presença de anião brometo
257	"	B2	Presença de anião cloreto
258	"	B3	Presença de anião tetrafluoroborato
259	"	B4	Presença de anião hexafluorofosfato
260	"	B5	Presença de anião trifluorometanosulfonato
261	"	B6	Presença de anião bis(trifluorometanosulfonil)imidato
262	"	B1Mc	B1 multiplicado por peso molecular do catião
263	"	B2Mc	B2 multiplicado por peso molecular do catião
264	"	B3Mc	B3 multiplicado por peso molecular do catião
265	"	B4Mc	B4 multiplicado por peso molecular do catião
266	"	B5Mc	B5 multiplicado por peso molecular do catião
267	"	B6Mc	B6 multiplicado por peso molecular do catião
268	"	B1Ma	B1 multiplicado por peso molecular do anião

Tabela II.1-3: Continuação da apresentação dos descritores base usados no estudo de modelação de pontos de fusão de sais de imidazólio.

Descritor número	Ião	Descritor	Descrição
269	Anião	B2Ma	B2 multiplicado por peso molecular do anião
270	"	B3Ma	B3 multiplicado por peso molecular do anião
271	"	B4Ma	B4 multiplicado por peso molecular do anião
272	"	B5Ma	B5 multiplicado por peso molecular do anião
273	"	B6Ma	B6 multiplicado por peso molecular do anião
274	"	B1Mca	B1 multiplicado pela diferença entre o peso molecular do catião e anião
275	"	B2Mca	B2 multiplicado pela diferença entre o peso molecular do catião e anião
276	"	B3Mca	B3 multiplicado pela diferença entre o peso molecular do catião e anião
277	"	B4Mca	B4 multiplicado pela diferença entre o peso molecular do catião e anião
278	"	B5Mca	B5 multiplicado pela diferença entre o peso molecular do catião e anião
279	"	B6Mca	B6 multiplicado pela diferença entre o peso molecular do catião e anião
280	"	Mca	diferença entre o peso molecular do catião e anião
281	"	absMca	valor absoluto da diferença entre o peso molecular do catião e anião

b) Conjunto de Sais de Imidazólio e Respectivas Classes Associadas ao Ponto de Fusão e Estado Físico à Temperatura Ambiente

Os 281 descritores usados neste estudo são os mesmos que foram utilizados no estudo da secção a). O modelo *Random Forest* baseou-se em todos os 281 descritores representados na **Tabela II.1**. Os modelos *Support Vector Machines* foram construídos usando subconjuntos do conjunto total de 71 descritores aniónicos e 210 descritores catiónicos. Os subconjuntos considerados consistiram nos descritores mais importantes na construção do modelo *Random Forest*. O subconjunto óptimo de descritores consistiu em 50 descritores (descriptor 15, 16, 25, 27, 28, 29, 30, 31, 35, 36, 37, 39, 44, 48, 50, 53, 55, 57, 61, 62, 64, 65, 66, 78, 93, 123, 129, 130, 131, 134, 168, 171, 181, 190, 196, 198, 205, 206, 225, 226, 228, 253, 262, 265, 266, 274, 277, 278, 280 e 281 – **Tabela II.1**) sendo aquele conjunto que proporcionava concordância mais elevada para conjunto de treino segundo validação cruzada (partindo conjunto de treino em 5 subconjuntos). Pela Árvore de Decisão foram seleccionados 10 descritores (descriptor 14, 20, 31, 49, 171, 190, 198, 211, 216 e 262 – **Tabela II.1**) do conjunto total de 281 descritores.

c) Conjunto de Sais de Imidazólio e Respectiva Densidade

Para a representação da estrutura de cada sal foram utilizados 246 descritores, 199 representando o catião e 47 representando o anião. Relativamente ao catião os descritores considerados incluem 128 descritores RDF considerando polarizabilidades atómicas e todos os hidrogénios presentes na molécula. Foram ainda determinados 26 descritores que representam a superfície do acessível do catião tendo em conta a electronegatividade sigma associada a cada átomo (descritores SURFACE). Adicionalmente, foram calculados 45 descritores XPET. Para representar o anião foram considerados 47 descritores aniónicos. Os descritores aniónicos baseiam-se em descritores binários (indicando a presença de determinado anião) e seus derivados multiplicando-os pelo peso molecular do catião, peso molecular do anião e diferença entre peso molecular de catião e anião. No caso do modelo *Random Forest* foram usados os 246 descritores moleculares representados na **Tabela II.2**. No caso dos modelos baseados em *Support Vector Machines* foram seleccionados diversos subconjuntos do conjunto total de 246 descritores moleculares, os subconjuntos considerados consistiam nos descritores mais importantes na construção do modelo *Random Forest*. O subconjunto óptimo de descritores estava composto por 60 descritores (descriptor 8, 9, 11, 14, 15, 16, 18, 21, 25, 26, 30, 31, 32, 39, 48, 89, 129, 130, 131, 157, 160, 161, 163, 170, 172, 193, 194, 195, 196, 200, 201, 202, 203, 204, 206, 207, 211, 212, 213, 215, 216, 218, 219,

221, 222, 223, 224, 225, 228, 229, 230, 231, 234, 235, 236, 237, 239, 240, 241 e 242 – **Tabela II.2**), aquele conjunto que proporcionava R^2 mais elevado para conjunto de treino segundo validação cruzada (partindo conjunto de treino em 3 subconjuntos), relativamente ao modelo baseado em árvore de decisão foram seleccionados por este algoritmo 3 descritores (descritor 9, 130 e 203 – **Tabela II.2**). Aparecem seguidamente os 246 descritores base considerados neste estudo (**Tabela II.2**):

Tabela II.2-1: 246 descritores base usados no estudo da densidade de sais de imidazólio.

Descritor número	Ião	Descritor	Descrição
1	Catião	APcH1	Descritor RDF considerando a polarizabilidade de átomos (sem H's) que distem entre si perto de 0,1 Å
2	APcH2 0,2 Å
3 - 127	APcH3 - APcH127	Descritores RDF considerando a polarizabilidade de átomos (sem H's) que distem entre si perto de 0,3 – 12,7 Å
128	APcH128 12,8 Å
129	EN1	Descritor SURFACE considerando pontos a distar entre si 0-1 Å
130	EN2 1-2 Å
131 - 153	EN3 – EN25	Descritores SURFACE considerando pontos a distar entre si 2-3 a 24-25 Å
154	EN26	Descritor SURFACE considerando pontos a distar entre si 25 - 26 Å
155	XP1	ATOMS – Número de átomos
156	XP2	BONDS – Número de ligações
157	XP3	MW – Massa molar
158	XP4	ALIPH_IND – Índice alifático
159	XP5	AROM_IND – Índice aromático
160	XP6	H_ACC – Potencial máximo de aceitador de hidrogénio
161	XP7	H_DON – Potencial máximo de doador de hidrogénio
162	XP8	MMPOL – Polarizabilidade
163	XP9	NH – Número de NH
164	XP10	NH2 – Número de NH ₂
165	XP11	HEAVY – Número de átomos pesados
166	XP12	OH – Número de grupos hidroxilo
167	XP13	NrO – Número de oxigénios
168	XP14	NrN – Número de azotos
169	XP15	QMIN – Carga mínima
170	XP16	QMAX – Carga máxima
171	XP17	MAX_CHAIN – Cadeia mais longa
172	XP18	N_CHAINS – Número de cadeias
173	XP19	DIST_CH[1]
174	XP20	DIST_CH[2]
175	XP21	DIST_CH[3]
176	XP22	DIST_CH[4]
177	XP23	DIST_CH[5]
178	XP24	DIST_CH[6]
179	XP25	DIST_CH[7]
180	XP26	DIST_CH[8]
181	XP27	DIST_CH[9]
182	XP28	DIST_CH[10]
183	XP29	DIST_CH[11]
184	XP30	DIST_CH[12]
185	XP31	DIST_CH_TOP[1]
186	XP32	DIST_CH_TOP[2]

Tabela II.2-2: 246 descritores base usados no estudo da densidade de sais de imidazólio.

Descritor número	Ião	Descritor	Descrição
187	Catão	XP33	DIST_CH_TOP[3]
188	*****	XP34	DIST_CH_TOP[4]
189	*****	XP35	DIST_CH_TOP[5]
190	*****	XP36	DIST_Q[1]
191	*****	XP37	DIST_Q[2]
192	*****	XP38	DIST_Q[3]
193	*****	XP39	DIST_Q[4]
194	*****	XP40	DIST_Q[5]
195	*****	XP41	DIST_Q[6]
196	*****	XP42	DIST_Q[7]
197	*****	XP43	DIST_Q[8]
198	*****	XP44	DIST_Q[9]
199	*****	XP45	DIST_Q[10]
200	Anião	An1	peso molecular de anião
201	*****	An2	diferença entre o peso molecular do catião e anião
202	*****	An3	valor absoluto da diferença entre o peso molecular do catião e anião
203	*****	An4	razão entre o peso molecular do anião e catião
204	*****	An5	razão entre o peso molecular do catião e anião
205	*****	An6	Presença de anião tetrafluoroborato
206	*****	An7	Presença de anião hexafluorofosfato
207	*****	An8	Presença de anião bis(trifluorometanosulfonil)imidato
208	*****	An9	Presença de anião trifluoro-metanosulfonato
209	*****	An10	Presença de anião trifluoroacetato
210	*****	An11	Presença de anião 2-hidroxi-propionato
211	*****	An12	An6 multiplicado pelo peso molecular do catião
212	*****	An13	An7 multiplicado pelo peso molecular do catião
213	*****	An14	An8 multiplicado pelo peso molecular do catião
214	*****	An15	An9 multiplicado pelo peso molecular do catião
215	*****	An16	An10 multiplicado pelo peso molecular do catião
216	*****	An17	An11 multiplicado pelo peso molecular do catião
217	*****	An18	An6 multiplicado pelo peso molecular do anião
218	*****	An19	An7 multiplicado pelo peso molecular do anião
219	*****	An20	An8 multiplicado pelo peso molecular do anião
220	*****	An21	An9 multiplicado pelo peso molecular do anião
221	*****	An22	An10 multiplicado pelo peso molecular do anião
222	*****	An23	An11 multiplicado pelo peso molecular do anião
223	*****	An24	An6 multiplicado pela diferença entre peso molecular do catião e anião
224	*****	An25	An7 multiplicado pela diferença entre peso molecular do catião e anião
225	*****	An26	An8 multiplicado pela diferença entre peso molecular do catião e anião
226	*****	An27	An9 multiplicado pela diferença entre peso molecular do catião e anião

Tabela II.2-3: 246 descritores base usados no estudo da densidade de sais de imidazólio.

Descritor número	Ião	Descritor	Descrição
227	Aniã	An28	An10 multiplicado pela diferença entre peso molecular do catião e aniã
228	An29	An11 multiplicado pela diferença entre peso molecular do catião e aniã
229	An30	An6 multiplicado pelo valor absoluto da diferença entre peso molecular do catião e aniã
230	An31	An7 multiplicado pelo valor absoluto da diferença entre peso molecular do catião e aniã
231	An32	An8 multiplicado pelo valor absoluto da diferença entre peso molecular do catião e aniã
232	An33	An9 multiplicado pelo valor absoluto da diferença entre peso molecular do catião e aniã
233	An34	An10 multiplicado pelo valor absoluto da diferença entre peso molecular do catião e aniã
234	An35	An11 multiplicado pelo valor absoluto da diferença entre peso molecular do catião e aniã
235	An36	An6 multiplicado pela razão entre peso molecular do aniã e catião
236	An37	An7 multiplicado pela razão entre peso molecular do aniã e catião
237	An38	An8 multiplicado pela razão entre peso molecular do aniã e catião
238	An39	An9 multiplicado pela razão entre peso molecular do aniã e catião
239	An40	An10 multiplicado pela razão entre peso molecular do aniã e catião
240	An41	An11 multiplicado pela razão entre peso molecular do aniã e catião
241	An42	An6 multiplicado pela razão entre peso molecular do catião e aniã
242	An43	An7 multiplicado pela razão entre peso molecular do catião e aniã
243	An44	An8 multiplicado pela razão entre peso molecular do catião e aniã
244	An45	An9 multiplicado pela razão entre peso molecular do catião e aniã
245	An46	An10 multiplicado pela razão entre peso molecular do catião e aniã
246	An47	An11 multiplicado pela razão entre peso molecular do catião e aniã

d) Conjunto de Sais de Guanidínio e Respetivo Ponto de Fusão

Representando a estrutura de cada catião deste conjunto de dados foram determinados 92 descritores moleculares. O programa Dragon 3.0¹⁰⁵ foi usado para determinar 52 descritores, incluindo 14 descritores constitucionais relacionados com o número de ligações (duplas, triplas, aromáticas) e número de átomos específicos (C, N, O, H), 4 descritores de grupos funcionais contabilizando o número de hidrogénios doadores e aceitadores e também relacionados com o grau de substituição do anel aromático, 33 descritores topológicos relacionados com flexibilidade e simetria estrutural, 1 descritor centrado num determinado fragmento atómico (representando o número de hidrogénios ligados a heteroátomo, correspondendo neste caso ao número de posições não substituídas no núcleo guanidínico). Considerando o programa PETRA 3.1 foram calculados 28 descritores adicionais de maneira a codificar o tamanho do catião, índices aromático e alifático e a distribuição de cadeias lineares. Finalmente 12 novos descritores foram determinados para codificar o tamanho dos grupos substituintes ligados ao núcleo guanidínico. Neste caso foram determinados seis descritores codificando o número de átomos que não sejam átomos de hidrogénio, um para cada posição do núcleo guanidínico (**Figura II.14**).

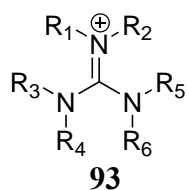


Figura II.14: Estrutura de um catião guanidínio evidenciando os seus seis grupos substituintes.

Para esse cálculo os átomos de azoto do núcleo de guanidínio foram ordenados de acordo com o número de átomos que não sejam átomos de hidrogénio associados aos dois grupos substituintes associados ao átomo de azoto a ser ordenado, posteriormente cada um dos dois substituintes desse átomo de azoto foram ordenados de acordo com o mesmo critério.

No caso do número de átomos não conseguir distinguir qual o grupo prioritário critérios como o peso molecular e tamanho da cadeia principal foram usados na decisão. Estes 92 descritores foram utilizados na construção de um modelo baseado numa rede neuronal de *Counter-Propagation*, e adicionalmente foi construído outro modelo baseado num conjunto de 5 redes de *Counter-Propagation*. Foi construído ainda um terceiro modelo baseado também num conjunto de 5 redes de *Counter-Propagation* usando um subconjunto de 16 descritores (descritor 18, 21, 25, 28, 31, 33, 36, 37, 43, 65, 69, 75, 82, 84, 85 e 87 – **Tabela II.3**) provenientes do conjunto de 92 descritores original. Seguidamente (**Tabela II.3**) aparecem discriminados os 92 descritores utilizados neste estudo.

Tabela II.3-1: 92 descritores base usados no estudo do ponto de fusão de sais de guanidínio.

Descritor número	Ião	Descritor	Descrição
1	Catião	ATOMS	Número de átomos
2	*****	BONDS	Número de ligações
3	*****	MW	Peso molecular do catião
4	*****	ALIPH_IND	Índice alifático
5	*****	AROM_IND	Índice aromático
6	*****	NH	Número de grupos NH
7	*****	NH2	Número de grupos NH ₂
8	*****	HEAVY	Número de átomos pesados
9	*****	OH	Número de grupos OH
10	*****	MAX_CHAIN	Cadeia mais longa
11	*****	N_CHAINS	Número de cadeias
12	*****	DIST_CH[1]	
13	*****	DIST_CH[2]	
14	*****	DIST_CH[3]	
15	*****	DIST_CH[4]	
16	*****	DIST_CH[5]	
17	*****	DIST_CH[6]	
18	*****	DIST_CH[7]	
19	*****	DIST_CH[8]	
20	*****	DIST_CH[9]	
21	*****	DIST_CH[10]	
22	*****	DIST_CH[11]	
23	*****	DIST_CH[12]	
24	*****	DIST_CH_TOP[1]	
25	*****	DIST_CH_TOP[2]	
26	*****	DIST_CH_TOP[3]	
27	*****	DIST_CH_TOP[4]	
28	*****	DIST_CH_TOP[5]	
29	*****	H-050	Número de hidrogénios ligados a heteroátomos
30	*****	nBM	Número de ligações múltiplas
31	*****	nCIC	Número de anéis
32	*****	RBN	Número de ligações rotáveis
33	*****	RBF	Fracção de ligações rotáveis
34	*****	nDB	Número de ligações duplas
35	*****	nTB	Número de ligações triplas
36	*****	nAB	Número de ligações aromáticas
37	*****	nH	Número de hidrogénios
38	*****	nC	Número de carbonos
39	*****	nN	Número de azotos
40	*****	nO	Número de oxigénios

Tabela II.3-2: 92 descritores base usados no estudo do ponto de fusão de sais de guanidínio.

Descritor número	Ião	Descritor	Descrição
41	Catião	nR05	Número de anéis de 5 átomos
42	nR06	Número de anéis de 6 átomos
43	nBnz	Número de anéis benzênicos
44	nCaH	Número de C(sp ²) aromáticos não substituídos
45	nCaR	Número de C(sp ²) aromáticos substituídos
46	nHDon	Número de átomos doadores em ligações por pontes de hidrogénio
47	nHAcc	Número de átomos aceitadores em ligações por pontes de hidrogénio
48	S0K	Índice de simetria de Kier
49	PHI	Índice Kier de flexibilidade
50	BLI	Índice Kier de semelhança a Benzeno
51	IC0	Índice de conteúdo de informação (ordem 0)
52	TIC0	Índice de conteúdo de informação total (ordem 0)
53	SIC0	Índice de conteúdo de informação estrutural (ordem 0)
54	CIC0	Índice de conteúdo de informação complementar (ordem 0)
55	BIC0	Índice de conteúdo de informação de ligação (ordem 0)
56	IC1	Índice de conteúdo de informação (ordem 1)
57	TIC1	Índice de conteúdo de informação total (ordem 1)
58	SIC1	Índice de conteúdo de informação estrutural (ordem 1)
59	CIC1	Índice de conteúdo de informação complementar (ordem 1)
60	BIC1	Índice de conteúdo de informação de ligação (ordem 1)
61	IC2	Índice de conteúdo de informação (ordem 2)
62	TIC2	Índice de conteúdo de informação total (ordem 2)
63	SIC2	Índice de conteúdo de informação estrutural (ordem 2)
64	CIC2	Índice de conteúdo de informação complementar (ordem 2)
65	BIC2	Índice de conteúdo de informação de ligação (ordem 2)
66	IC3	Índice de conteúdo de informação (ordem 3)
67	TIC3	Índice de conteúdo de informação total (ordem 3)
68	SIC3	Índice de conteúdo de informação estrutural (ordem 3)
69	CIC3	Índice de conteúdo de informação complementar (ordem 3)
70	BIC3	Índice de conteúdo de informação de ligação (ordem 3)

Tabela II.3-3: 92 descritores base usados no estudo do ponto de fusão de sais de guanidínio.

Descritor número	Ião	Descritor	Descrição
71	Catião	IC4	Índice de conteúdo de informação (ordem 4)
72	TIC4	Índice de conteúdo de informação total (ordem 4)
73	SIC4	Índice de conteúdo de informação estrutural (ordem 4)
74	CIC4	Índice de conteúdo de informação complementar (ordem 4)
75	BIC4	Índice de conteúdo de informação de ligação (ordem 4)
76	IC5	Índice de conteúdo de informação (ordem 5)
77	TIC5	Índice de conteúdo de informação total (ordem 5)
78	SIC5	Índice de conteúdo de informação estrutural (ordem 5)
79	CIC5	Índice de conteúdo de informação complementar (ordem 5)
80	BIC5	Índice de conteúdo de informação de ligação (ordem 5)
81	#1	Número de átomos (não hidrogénios) ligados à posição 1 do núcleo de guanidínio
82	#2	Número de átomos (não hidrogénios) ligados à posição 2 do núcleo de guanidínio
83	#3	Número de átomos (não hidrogénios) ligados à posição 3 do núcleo de guanidínio
84	#4	Número de átomos (não hidrogénios) ligados à posição 4 do núcleo de guanidínio
85	#5	Número de átomos (não hidrogénios) ligados à posição 5 do núcleo de guanidínio
86	#6	Número de átomos (não hidrogénios) ligados à posição 6 do núcleo de guanidínio
87	cadeia(max-min)	Diferença entre a maior e a menor cadeias ligadas a uma das 6 posições no núcleo de guanidínio
88	sum(d-AlkN)	Soma das diferenças entre o tamanho de cadeias ligadas ao mesmo azoto do núcleo de guanidínio
89	nMax	Número de átomos (não hidrogénios) na cadeia contendo número mais elevado de átomos
90	nMin	Número de átomos (não hidrogénios) na cadeia contendo número mais baixo de átomos
91	#Max	Número de cadeias contendo número máximo de átomos (não hidrogénios).
92	#Min	Número de cadeias contendo número mínimo de átomos (não hidrogénios).

II-2.1.2: Normalização de Dados

Em cada um dos conjuntos de dados considerados neste capítulo, cada descriptor foi normalizado linearmente de forma a ter um valor compreendido entre 0 e 1, a expressão utilizada na normalização foi a seguinte:

$$descriptor_normalizado = \frac{descriptor_original - \min}{\max - \min}$$

Equação II.40

Nesta equação *max* e *min* representam, respectivamente, o valor máximo e mínimo de um determinado descriptor, entre os valores de todos os objectos considerados no estudo. Esta normalização foi utilizada quando o método de aprendizagem automático utilizado na construção do modelo era uma rede de *Counter-Propagation* já que estas assim o exigem para que todos os descritores tenham a mesma importância à partida. Quando outros métodos de aprendizagem foram utilizados, como o caso dos métodos *Random Forest* e *Árvore de Decisão* não foi necessário normalizar os descritores, e no caso dos modelos *Support Vector Machines* a normalização é interna (normalização Z).

II-2.1.3: Representação de sais

Os sais estudados ao longo deste capítulo possuem mais que uma estrutura de ressonância, e diferentes estruturas de ressonância resultam em descritores diferentes. Para resolver essa questão cada sal foi representado por todas as suas estruturas de ressonância. No caso dos sais de imidazólio cada sal foi representado por duas estruturas de ressonância (**Figura II.15**):

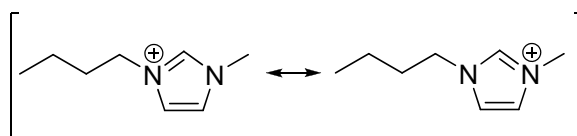


Figura II.15: Duas estruturas de ressonância de um catião imidazólio com a carga positiva a estar centrada nos dois átomos de azoto.

Cada estrutura de ressonância conta como um objecto, sendo calculados descritores moleculares para todos os objectos. Este procedimento foi usado na modelação do ponto de fusão de sais de imidazólio **a**), na classificação de sais de imidazólio segundo o seu ponto de fusão e estado físico **b**), e modelação da sua densidade **c**). No caso de sais de guanidínio, estudo **d**), seguiu-se a mesma estratégia sendo cada sal representado por três estruturas de ressonância (**Figura II.16**):

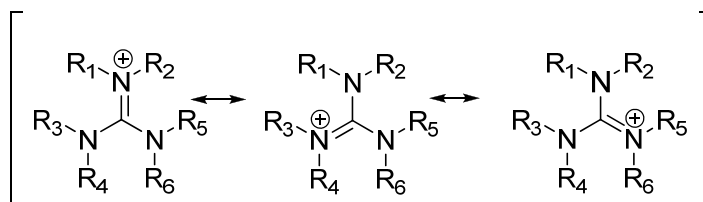


Figura II.16: As três estruturas de ressonância mais relevantes num catião baseado na unidade guanidínio com a carga positiva a estar centrada nos átomos de azoto.

Assim, cada estrutura de ressonância representa um objecto, sendo calculados descritores para todos os objectos.

II-2.1.4: Características Específicas dos conjuntos de dados

Para poder avaliar a capacidade de previsão para novos casos um modelo necessita de ser validado por um conjunto de objectos que não tenha sido utilizado para a sua obtenção. Daí a necessidade de dividir os conjuntos de dados num conjunto de treino, para construção do modelo, e um conjunto de teste para a sua validação. Adicionalmente, em cada conjunto de dados existem famílias aniónicas e classes de propriedades físico-químicas, sendo descritas todas estas características para cada um dos quatro conjuntos de dados:

- a) **Conjunto de Sais de Imidazólio e Respectivo Ponto de Fusão:** Dividido num conjunto de treino de 308 objectos (154 sais) e conjunto de teste de 158 objectos (79 sais).

Tabela II.4: Composição dos conjuntos de treino e teste ao nível de distribuição de aniões e valores máximos e mínimos de ponto de fusão.

Anião	Treino			Teste		
	Nº de objectos total	Valor máximo de ponto de fusão (°C)	Valor mínimo de ponto de fusão (°C)	Nº de objectos total	Valor máximo de ponto de fusão (°C)	Valor mínimo de ponto de fusão (°C)
BF₄⁻	38	85	-5	12	71	22
Br⁻	124	370	5,5	98	283	12
Cl⁻	40	278	38	14	178	59,5
(CF₃SO₂)₂N⁻	28	98	-22	10	86	9
PF₆⁻	52	196	5	18	116	30
CF₃SO₃⁻	26	191	-10,5	6	109	8
Total	308	370	-22	158	283	8

- b) **Conjunto de Sais de Imidazólio e Respectivas Classes Associadas ao Ponto de Fusão e Estado Físico à Temperatura Ambiente:** Dividido num conjunto de treino de 436 objectos (218 sais) e num conjunto de teste de 218 objectos (109 sais).

Tabela II.5: Composição do conjunto de treino e conjunto de teste ao nível de aniões e classes.

Anião	Treino				Teste			
	Nº de objectos total	Nº de objectos classe a	Nº de objectos classe b	Nº de objectos classe c	Nº de objectos total	Nº de objectos classe a	Nº de objectos classe b	Nº de objectos classe c
BF_4^-	82	54	28	0	32	22	10	0
Br^-	126	6	14	106	100	4	12	84
Cl^-	42	2	24	16	16	2	6	8
$(\text{CF}_3\text{SO}_2)_2\text{N}^-$	68	52	16	0	30	24	6	0
PF_6^-	86	36	38	12	30	12	14	4
CF_3SO_3^-	32	12	14	6	10	6	2	2
Total	436	162	134	140	218	70	50	98

Classe a – sais líquidos a partir de temperaturas até 25°C, Classe b – sais com ponto de fusão compreendido entre 25 e 100°C, Classe c – sais com ponto de fusão superior a 100°C.

- c) **Conjunto de Sais de Imidazólio e Respectiva Densidade:** Dividido num conjunto de treino de 136 objectos (68 sais) e conjunto de teste de 68 objectos (34 sais).

Tabela II.6: Distribuição de objectos em conjuntos de treino e teste segundo anião e gamas de densidade.

Anião	Treino			Teste		
	Nº de objectos total	Valor máximo de densidade (g/cm^3)	Valor mínimo densidade (g/cm^3)	Nº de objectos total	Valor máximo de densidade (g/cm^3)	Valor mínimo densidade (g/cm^3)
BF_4^-	36	1,34	0,99	18	1,205	1,02
PF_6^-	22	1,41	1,09	10	1,335	1,125
$(\text{CF}_3\text{SO}_2)_2\text{N}^-$	40	2,8	1,205	18	2,35	1,249
CF_3SO_3^-	6	1,39	1,24	6	1,364	1,27
CF_3CO_2^-	6	1,29	1,183	2	1,25	1,25
$\text{CH}_3\text{C}(\text{OH})\text{CO}_2^-$	26	1,1081	0,9591	14	1,1231	0,9793
Total	136	2,8	0,9591	68	2,35	0,9793

- d) **Conjunto de Sais de Guanidínio e Respectivo Ponto de Fusão:** Dividido num conjunto de treino de 243 objectos (81 sais) e conjunto de teste de 60 objectos (20 sais).

Tabela II.7: Distribuição de objectos por conjunto de treino e teste segundo anião e gamas de densidade.

Anião	Treino			Teste		
	Nº de objectos total	Valor máximo de ponto de fusão (°C)	Valor mínimo de ponto de fusão (°C)	Nº de objectos total	Valor máximo de ponto de fusão (°C)	Valor mínimo de ponto de fusão (°C)
I ⁻	78	300	46	12	180	64
Br ⁻	27	214	93	6	213	94
Cl ⁻	93	293	-76	24	182	68
BPh ₄ ⁻	45	322	127	18	274	132
Total	243	322	-76	60	274	64

II-2.2: Modelos Construídos

Considerando os conjuntos de dados descritos na secção anterior, foram construídos 4 tipos de modelos: **a)** modelos para previsão do ponto de fusão de sais de imidazólio, **b)** modelos para classificação de sais de imidazólio segundo o seu estado físico/ponto de fusão, **c)** modelos para previsão da densidade de sais de imidazólio e **d)** modelos para previsão do ponto de fusão de sais de guanidínio. Na construção dos modelos baseados no algoritmo *Random Forest* foi usado o programa R versão 2.4.0 usando a *library randomForest*. Os modelos baseados no algoritmo da Árvore de Decisão foram construídos usando o programa R¹⁰⁶ versão 2.4.0 com a *library rpart*. Por sua vez os modelos baseados no algoritmo *Support Vector Machines* foram construídos usando o programa R versão 2.4.0 com a *library kernlab*. Relativamente aos modelos baseados em redes neuronais de *Counter-Propagation* foi usada aplicação Java desenvolvida no grupo e denominada JATOON (Java Tools for Neural Networks).^{107,108} Vão ser descritos seguidamente os parâmetros utilizados em cada modelo.

II-2.2.1 - Estudo A: Modelação de Pontos de Fusão de Sais de Imidazólio

Foram construídos diversos modelos para descrever e prever o ponto de fusão de sais de imidazólio em que foram usados diversos métodos de aprendizagem automática - *Random Forest*, *Support Vector Machines* e Árvore de Decisão. Para comparação dos diferentes modelos foi usado o método da validação cruzada dividindo o conjunto de treino em 5 subconjuntos. Quatro dos subconjuntos foram usados no treino de um modelo (com as mesmas características dos modelos descritos nas secções seguintes) e esse modelo foi usado na obtenção de previsão para o subconjunto restante. Este procedimento repetiu-se por mais 4 vezes até todos os subconjuntos terem sido usados na obtenção de previsões. As duas estruturas de ressonância a representar um determinado sal

pertenceram sempre ao mesmo subconjunto. O modelo que apresenta os melhores parâmetros em validação cruzada foi analisado mais detalhadamente sendo validado adicionalmente por um conjunto independente de teste. Seguem-se as características de cada modelo estudado.

II-2.2.1.1: Modelo *Random Forest*

Foram construídos quatro modelos *Random Forest* (ver Capítulo I) com os parâmetros *ntree* (número de árvores) = 1000, e *mtry* (número de descritores testado em cada nó) tomando valores 10, 25, 50 e 100. Finalmente entre os quatro modelos construídos foi seleccionado aquele cujo valor de *mtry* é 25 (*ntree* = 1000) sendo aquele que proporcionou valor de R^2 mais elevado em validação cruzada partindo o conjunto de treino em cinco subconjuntos, em que as duas estruturas de ressonância a representar um determinado sal vão pertencer ao mesmo subconjunto.

II-2.2.1.2: Modelos *Support Vector Machines*

Foi construída uma série de modelos *Support Vector Machines* variando o número de descritores utilizados (os mais importantes segundo modelo *Random Forest*) e os parâmetros C e ν . O critério de escolha do melhor modelo assentou no valor R^2 para o conjunto de treino com validação cruzada (c.v. 5x), em que as duas estruturas de ressonância a representar um determinado sal pertenceram ao mesmo subconjunto.

O modelo seleccionado foi construído com os 30 descritores mais importantes seleccionados pelo modelo das *Random Forest*. Por sua vez o parâmetro ν associado a este modelo tem o valor 0,032 e C apresenta 32 como valor. O parâmetro ε tem o valor 0,1. O modelo construído é do tipo ν -svr, a função Kernel utilizada é Gaussiana cujo parâmetro sigma tem valor $\sigma = 0,0277553773798545$. O modelo assenta em 93 *Support Vectors*.

II-2.2.1.3: Modelo Árvore de Decisão

Foi construído um modelo Árvore de Decisão usando parâmetros de referência. O modelo final está composto por 12 regras lógicas (11 descritores).

II-2.2.2 - Estudo B: Modelos para Classificação de Sais de Imidazólio segundo Estado Físico/Ponto de Fusão

À semelhança do estudo II-2.2.1, foram construídos diferentes modelos usando métodos de aprendizagem automática, *Random Forest*, *Support Vector Machines* e Árvore de Decisão. Os diferentes modelos foram comparados recorrendo a validação cruzada (c.v.) dos objectos do conjunto de treino.

II-2.2.2.1: Modelo *Random Forest*

Foram construídos quatro modelos *Random Forest* (ver Capítulo I) com os parâmetros, *ntree* (número de árvores) = 1000, e *mtry* (número de descritores testado em cada nó) a variar: *mtry* = 10, 25, 50 e 100. Ainda que a variação de resultados com o parâmetro *mtry* seja pouco significativa, entre os quatro modelos construídos foi seleccionado aquele cujo valor de *mtry* é 25 (*ntree* = 1000) sendo aquele que proporcionou valor de concordância mais elevado em validação cruzada partindo o conjunto de treino em cinco subconjuntos, em que as duas estruturas de ressonância a representar um determinado sal vão pertencer ao mesmo subconjunto.

II-2.2.2.2: Modelos *Support Vector Machines*

Foram construídos uma série de modelos *Support Vector Machines* de forma a encontrar o modelo que melhor classificava sais de imidazólio segundo o seu estado físico/ponto de fusão. O procedimento utilizado é semelhante ao utilizado em II-2.2.1.2 no estudo de regressão, neste caso concreto o modelo seleccionado foi aquele cujo resultado na validação cruzada (c.v. 5x) proporcionava concordância mais elevada. O modelo seleccionado baseia-se em 50 descritores, os mais importantes (segundo modelo *Random Forest*), tendo o parâmetro *C* de complexidade o valor 8. O modelo construído é do tipo *C-svc* e a função *Kernel* é Gaussiana com parâmetro sigma $\sigma = 0,0194023407952181$. Este modelo assenta em 205 *Support Vectors*.

II-2.2.2.3: Modelo Árvore de Decisão

Foi construído um modelo Árvore de Decisão usando parâmetros de referência. O modelo final está composto por 10 regras lógicas (10 descritores).

II-2.2.3 - Estudo C: Modelação da Densidade de Sais de Imidazólio

Foram construídos diversos modelos de descrição e previsão de densidade de sais de imidazólio, recorrendo a métodos de aprendizagem automática descritos nas secções seguintes. Os diferentes modelos obtidos foram comparados usando o método da validação cruzada, tal como nas secções II-2.2.1 e II-2.2.2, a única diferença consiste no facto de se ter partido o conjunto de treino em 3 subconjuntos. O modelo que apresentava melhor capacidade de previsão foi adicionalmente validado através das previsões de densidade para um conjunto independente de teste.

II-2.2.3.1: Modelo *Random Forest*

Foram construídos quatro modelos *Random Forest* (ver Capítulo I) com os parâmetros, *ntree* (número de árvores) = 1000, e *mtry* (número de descritores testado em cada nó) a variar: *mtry* = 10, 25, 50 e 100. Finalmente entre os quatro modelos construídos foi seleccionado aquele cujo valor de *mtry* é 10 (*ntree* = 1000) sendo aquele que proporcionou valor de R^2 mais elevado em validação cruzada partindo o conjunto de treino em três subconjuntos, em que as duas estruturas de ressonância a representar um determinado sal vão pertencer ao mesmo subconjunto.

II-2.2.3.2: Modelos *Support Vector Machines*

À semelhança da secção II-2.2.1.2. e II-2.2.2.2., foi construída uma série de modelos *Support Vector Machines*, em que se variava o número de descritores utilizados (os mais importantes segundo modelo *Random Forest*), tal como o parâmetro ν . Posteriormente foi seleccionado o modelo que permitia a obtenção de R^2 mais elevados para conjunto de treino com validação cruzada (c.v. 3x), em que as duas estruturas de ressonância a representar um determinado sal vão pertencer ao mesmo subconjunto. O modelo seleccionado foi construído com os 60 descritores mais importantes seleccionados pelo modelo das *Random Forest*. Por sua vez o parâmetro ν associado a este modelo tem o valor 0,016, por sua vez o parâmetro C seleccionado tem como valor 16, o valor de ε é 0,1. O modelo construído é do tipo ν -svr, a função Kernel utilizada é Gaussiana cujo parâmetro sigma tem como valor $\sigma = 0,0176621355057598$. O modelo assenta em 66 *Support Vectors*.

II-2.2.3.3: Modelo Árvore de Decisão

Foi construído um modelo Árvore de Decisão usando parâmetros de referência. O modelo final está composto por 4 regras lógicas (4 descritores).

II-2.2.4 - Estudo D: Modelação do Ponto de Fusão de Sais de Guanidínio

No estudo do ponto de fusão de sais de guanidínio foram treinados uma série de modelos baseados em redes neuronais de *Counter-Propagation*; foi construído um modelo 1) baseado numa rede neuronal de *Counter-Propagation*. Foi ainda construído um modelo 2) baseado num *ensemble* de 5 redes neuronais de *Counter-Propagation*. Finalmente foi construído um modelo 3) baseado numa rede neuronal de *Counter-Propagation* em que houve selecção prévia de descritores baseada no mapeamento de descritores como objectos em rede neuronal de *Kohonen*. À semelhança das redes neuronais de *Counter-Propagation*, as redes neuronais de *Kohonen* foram treinadas e usadas através do programa JATOON.

II-2.2.4.1: Modelo Baseado numa Rede de *Counter-Propagation*

O modelo treinado é uma rede neuronal de *Counter-Propagation* de dimensões 18x18, com 92 camadas de *input* (92 descritores moleculares) e 4 camadas de *output* (cada camada corresponde a um anião diferente). Este modelo foi construído usando *initial learning span* = 8 e número de *epochs* = 100.

Validação de Modelo por Validação Cruzada

O modelo construído foi validado através de *Validação cruzada* dividindo o conjunto de treino em 5 subgrupos, como explicado anteriormente. Neste procedimento, os objectos do subgrupo usado para previsão que tivessem ponto de fusão superior ao máximo ou inferior ao mínimo dos outros 4 subconjuntos foram retirados. Neste ensaio as três estruturas de ressonância a representar um determinado sal pertenceram ao mesmo subconjunto. As redes neuronais de *Counter-Propagation* usadas para este fim foram treinadas com os mesmos parâmetros usados no estudo original (II-2.2.4.1 - Modelo baseado numa rede de *Counter-Propagation*).

Validação de Modelo através de Distribuição Aleatória de Pontos de Fusão por Objectos do Conjunto de Treino

O modelo construído foi adicionalmente validado através de distribuição aleatória dos pontos de fusão do conjunto de treino de 243 objectos (81 sais) – *y-randomization*. Posteriormente, o conjunto de treino assim obtido foi usado no treino de uma rede de *Counter-Propagation*, que foi usada para obter previsões para o conjunto de teste de 60 objectos (20 sais). Este procedimento repetiu-se por 4 ensaios adicionais, em cada ensaio havia uma distribuição aleatória diferente. As redes neuronais de *Counter-Propagation* usadas para este fim foram treinadas com os mesmos parâmetros usados no estudo original (II-2.2.4.1 - Modelo baseado numa rede de *Counter-Propagation*).

II-2.2.4.2: Modelo baseado num Conjunto de 5 redes de *Counter-Propagation*

O modelo treinado é um conjunto de 5 redes neuronais de *Counter-Propagation*, uma das redes é a mesma que foi utilizada no estudo original (II-2.2.4.1 - Modelo baseado numa rede de *Counter-Propagation*), as restantes 4 redes foram treinadas usando os mesmos parâmetros e possuindo as mesmas características que a rede descrita na secção anterior: dimensões 18x18, 92 camadas de *input*, 4 camadas de *output*, *initial learning span* 8 e número de *epochs* 100. O valor de

ponto de fusão previsto para um determinado objecto é a média em relação aos valores de ponto de fusão apresentados por cada uma das 5 redes *CPG-NN*.

II-2.2.4.3: Modelo baseado em Rede de *Counter-Propagation* Treinada com Conjunto Seleccionado de Descritores

Partindo dos 92 descritores usados no estudo original (II-2.2.4.1 - **Modelo baseado numa rede de *Counter-Propagation***) foram seleccionados 16 para obtenção de um modelo baseado numa rede de *Counter-Propagation*. Para a selecção dos 16 descritores a matriz contendo os objectos do conjunto treino foi transposta para que cada um dos descritores e pontos de fusão passassem a ser um objecto, a matriz assim transposta foi usada para treinar uma rede neuronal de *Kohonen*¹⁰⁹ 10x10, com 243 camadas de *input*. Foi usado *initial learning span* = 5 e número de *epochs* = 100. A rede de *Kohonen* assim obtida foi usada para mapear os 92 descritores e 4 pontos de fusão (correspondendo cada ponto de fusão a um anião diferente), que através da transposição da matriz do conjunto de treino passaram a ser objectos (96 objectos). Nesse mapa retirou-se um descritor por cada neurónio ocupado, sendo seleccionados 16 descritores. Representado pelos 16 descritores seleccionados, o conjunto de treino original foi usado na obtenção de um modelo baseado num conjunto de 5 redes de *Counter-Propagation*. Cada uma das 5 redes *CPG-NN* tinha dimensão 18x18, 16 camadas de *input* (16 descritores moleculares) e 4 camadas de *output* (uma camada para cada anião diferente). Ao nível dos parâmetros de treino o *initial learning span* teve valor 8 e número de *epochs* 100. O valor de ponto de fusão previsto para um determinado objecto é a média em relação aos valores de ponto de fusão apresentados por cada uma das 5 redes *CPG-NN*. O modelo foi validado por validação cruzada usando procedimento apresentado na secção **Validação de modelo por Validação Cruzada** em II-2.2.4.1.

II-2.3: Cálculo do Valor de Densidade e Ponto de Fusão Previsto por Modelos Computacionais

Como foi referido na secção II-2.1.3 cada sal estudado foi representado por uma série de estruturas de ressonância, 2 no caso de sais de imidazólio e 3 no caso de sais de guanidínio. Cada estrutura de ressonância foi considerada como um objecto independente na construção de determinado modelo e obtenção de previsões. Nos estudos **a)**, **c)** e **d)** a previsão associada a determinado sal é a média da propriedade em estudo tendo em conta todas as estruturas de ressonância que representam esse sal. No modelo de classificação **b)** a classe prevista para determinado sal é a classe que apresenta unanimidade para as 2 estruturas de ressonância.

II-2.4: Medidas de Qualidade dos Modelos de Previsão

II-2.4 - Estudo A: Modelação de Pontos de Fusão de Sais de Imidazólio

Os diversos modelos construídos foram comparados de acordo com os seguintes parâmetros:

Erro médio absoluto (MAE): Tal como o próprio nome indica, é o valor médio do desvio em valor absoluto definido por:

$$MAE = \frac{\sum_{i=1}^N |y_{prev} - y|}{N}$$

Equação II.41

em que $|y_{prev} - y|$ é o valor absoluto do desvio entre o valor previsto y_{prev} e o valor verdadeiro y . Esta medida é obtida para um conjunto de N objectos.

Pearson, coeficiente de correlação (R): Medida estatística que visa avaliar a força e a direcção de uma relação linear entre duas variáveis, definida por:

$$R = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

Equação II.42

Em que x e y são os valores dessas variáveis, e \bar{x} e \bar{y} as suas respectivas médias.

Raiz do erro médio quadrático, (RMS): medida dos desvios entre valores experimentais de cada objecto e_i e respectivos valores previstos p_i , definida por:

$$RMS = \sqrt{\frac{\sum_{i=1}^N (e_i - p_i)^2}{N}}$$

Equação II.43

em que N é o número total de objectos.

Spearman, Coeficiente (SP): É uma medida de correlação dos rankings dos valores associados a dois tipos de variáveis, e aparece definido matematicamente como:

$$1 - \frac{6 \times \left(\sum_i^N (\text{Ranking}(y_i) - \text{Ranking}(x_i))^2 \right)}{N \times (N^2 - 1)}$$

Equação II.44

Em que $\text{Ranking}(y_i)$ e $\text{Ranking}(x_i)$ são os rankings associados ao valor i das variáveis x e y , por sua vez N é o número de objectos ou valores presentes em cada variável, neste caso parte-se do pressuposto que existe o mesmo número de valores em cada variável.

II-2.4 - Estudo B: Modelos para Classificação de Sais de Imidazólio segundo Estado Físico/Ponto de Fusão

Nos modelos de classificação construídos foram usadas uma série de medidas de comparação:

concordância – É a percentagem de previsões correctas entre todos os objectos submetidos a determinado modelo específico

k - É a percentagem de previsões correctas entre todos os objectos submetidos a determinado modelo descontando a percentagem de previsões correctas obtidas de forma aleatória. Este parâmetro aparece definido matematicamente:

$$k = \frac{\left(\frac{1}{N} \right) \sum_i n_{ii} - \left(\frac{1}{N^2} \right) \sum_i n_{.i} n_{i.}}{1 - \left(\frac{1}{N^2} \right) \sum_i n_{.i} n_{i.}}$$

Equação II.45

Em que N é o número total de objectos (sais), com n_{ii} a representar os membros da diagonal da tabela de contingência (**Tabela II.8** - número de previsões correctas associadas a cada classe) com $n_{.i} n_{i.}$ a representar o produto entre o número total de elementos a uma determinada linha e o número total de elementos de uma coluna com o mesmo índice:

Tabela II.8: Exemplo de uma tabela de contingência para 3 classes distintas.

		Classe experimental			Total-linha
		a	b	c	
Classe prevista	a	36	7	1	44
	b	5	16	4	25
	c	1	4	47	52
	Total-coluna	42	27	52	121

A diagonal principal está assinalada a negrito e representa os objectos correctamente classificados. Fora da diagonal principal estão objectos em que não há correspondência entre a classe experimental e a classe prevista.

λ_B – Permite estabelecer a capacidade de previsão de determinado modelo descontando o efeito de uma distribuição assimétrica de elementos de uma dada classe, esta medida aparece definida matematicamente por:

$$\lambda_B = \frac{\sum_i \max_j(n_{ij}) - \max_j(n_{.j})}{N - \max_j(n_{.j})}$$

Equação II.46

Em que $\max_j(n_{ij})$ é o valor máximo associado a uma coluna j , por sua vez $\max_j(n_{.j})$ é o valor máximo do número total de elementos associado a determinada coluna entre o conjunto total de colunas, N é o número total de elementos do conjunto de objectos em estudo.

predictibilidade associada a uma classe, é a percentagem de acertos entre todos os objectos previstos com determinada classe.

sensibilidade associada a uma classe, é a percentagem de objectos pertencendo à mesma classe experimental que foram correctamente previstos.

II-2.4 - Estudo C: Modelação da Densidade de Sais de Imidazólio

Foram considerados os mesmos parâmetros usados no estudo a) de regressão.

II-2.4 - Estudo D: Modelação do Ponto de Fusão de Sais de Guanidínio

Foram considerados os mesmos parâmetros usados no estudo a) de regressão.

II-3: Resultados e Discussão

Ao longo deste capítulo são descritos seis estudos distintos tanto no objecto de estudo como na abordagem seguida na sua execução, desde abordagens complementares em Químio-informática/Síntese Orgânica até abordagens exclusivamente Químio-Informáticas e exclusivamente em Síntese Orgânica. Os seis estudos são: **a)** modelação de ponto de fusão de sais de imidazólio, **b)** classificação de sais de imidazólio segundo ponto de fusão/estado físico, **c)** modelação da densidade de sais de imidazólio, **d)** modelação do ponto de fusão de sais de guanidínio, **e)** síntese de novos materiais densos baseados em sais de guanidínio e **f)** síntese de sais de guanidínio baseados em aniões dicianamida, tiocianato e para-tolueno-sulfonato.

II-3.1 - Estudo A: Modelação do Ponto de Fusão de Sais de Imidazólio

No estudo descrito nesta secção pretende-se modelar o ponto de fusão de sais de imidazólio, para esse fim foram construídos e comparados três modelos baseados em métodos de aprendizagem automática distintos. Os três modelos de previsão de ponto de fusão foram comparados na capacidade de ajuste aos dados experimentais do conjunto de treino, sendo seleccionado para posterior estudo aquele modelo que apresentasse melhor capacidade de previsão em validação cruzada (*c.v.*), esse modelo foi validado com conjunto independente de compostos. Os três modelos construídos apresentam uma série de características particulares e que importa aqui apresentar.

Tabela II.9: Características dos modelos construídos para modelar o ponto de fusão de sais de imidazólio.

Método de aprendizagem automático	Número de descritores	Tipo de descritores	Normalização dos descritores	Parâmetros associados ao método
<i>Random Forest</i>	281	Catião e Anião	Não	1000 árvores, <i>mtry</i> = 25
Árvore de Decisão	11	Catião e Anião	Não	referência
<i>Support Vector Machines</i>	30	Catião e Anião	Interna (normalização Z)	$\nu = 0,032$ $C=32$

Através da análise à **Tabela II.9** pode-se observar que os três modelos construídos incluem tanto descritores catiónicos como aniónicos, o modelo *Random Forest* assenta em 281 descritores moleculares, os descritores não foram normalizados pois o modelo é construído de forma sequencial. Relativamente ao modelo Árvore de Decisão foram usados 11 descritores que não foram normalizados dado o mecanismo sequencial de construção deste modelo. No caso do modelo *Support*

Vector Machines foram usados 30 descritores neste estudo de regressão, os descritores mais importantes para o modelo *Random Forest*

Os modelos descritos na **Tabela II.9** consideram sais de imidazólio contendo como anião o brometo (Br^-), o cloreto (Cl^-), o tetrafluoroborato (BF_4^-), o hexafluorofosfato (PF_6^-), o trifluorometanosulfonato (CF_3SO_3^-) e o bis(trifluorometanosulfonil)imidato ($\text{N}(\text{SO}_2\text{CF}_3)_2^-$). Os modelos assim construídos foram comparados entre si, tanto na capacidade de modelar os pontos de fusão experimentais dos objectos do conjunto de treino, como na capacidade de previsão através de validação cruzada (ver glossário) dividindo o conjunto de treino em 5 subconjuntos, os resultados obtidos aparecem condensados na **Tabela II.10**:

Tabela II.10: Comparação, entre modelos RF, SVM e AD aplicados ao conjunto de treino com e sem validação cruzada 5x.

		Conjunto de Treino			Validação Cruzada (5x)		
		RF	SVM	AD	RF	SVM	AD
Geral	R ²	0,985	0,931	0,859	0,670	0,664	0,433
	RMS (°C)	10,287	19,705	27,805	42,508	42,895	57,364
	SP	0,991	0,954	0,883	0,835	0,843	0,683
	MAE (°C)	6,831	17,265	20,521	28,840	29,879	40,680
	n sais	154					
Br ⁻	R ²	0,986	0,909	0,735	0,341	0,373	0,110
	RMS (°C)	11,872	21,077	33,839	54,217	53,450	70,029
	SP	0,994	0,992	0,804	0,572	0,580	0,398
	MAE (°C)	8,905	19,441	25,247	40,630	39,176	51,282
	n sais	62					
Cl ⁻	R ²	0,962	0,917	0,863	0,420	0,256	0,093
	RMS (°C)	12,544	20,983	21,665	43,062	47,196	60,410
	SP	0,977	0,998	0,707	0,452	0,470	0,193
	MAE (°C)	7,621	18,282	17,688	30,895	32,972	45,052
	n sais	20					
BF ₄ ⁻	R ²	0,986	0,546	0,700	0,692	0,499	0,319
	RMS (°C)	3,879	17,455	14,198	14,978	18,290	34,188
	SP	0,996	0,799	0,378	0,836	0,717	0,723
	MAE (°C)	2,776	13,979	11,403	11,225	13,874	22,083
	n sais	19					
PF ₆ ⁻	R ²	0,978	0,831	0,728	0,473	0,607	0,148
	RMS (°C)	7,533	15,973	21,192	27,040	25,425	40,555
	SP	0,985	0,862	0,614	0,851	0,865	0,440
	MAE (°C)	3,439	13,137	17,082	13,632	16,933	30,327
	n sais	26					
CF ₃ SO ₃ ⁻	R ²	0,984	0,862	0,661	0,259	0,395	0,028
	RMS (°C)	11,816	21,705	31,815	46,663	42,602	62,870
	SP	0,973	0,929	0,558	0,610	0,687	0,209
	MAE (°C)	9,937	19,621	24,654	37,320	34,275	49,905
	n sais	13					
N(SO ₂ CF ₃) ₂ ⁻	R ²	0,958	0,792	0,471	0,534	0,772	0,08
	RMS (°C)	7,721	18,715	21,655	22,923	29,706	30,899
	SP	0,901	0,589	0,541	0,708	0,712	0,411
	MAE (°C)	5,435	16,119	18,564	17,965	23,610	23,380
	n sais	14					

Tanto ao nível geral como para cada tipo de ânion de cada conjunto de dados aparecem discriminadas medidas de qualidade dos modelos: R² – Quadrado do coeficiente de correlação de Pearson. RMS - Raiz do erro médio quadrático . SP – Coeficiente de Spearman. MAE – Erro médio absoluto.

Olhando para a **Tabela II.10** é possível constatar a comparação entre três modelos de regressão de ponto de fusão, *Random Forest* (RF), *Support Vector Machines* (SVM) e *Árvore de*

Decisão (AD). As medidas que servem de comparação são o quadrado do coeficiente de correlação Pearson (R^2), a raiz do erro médio quadrático (RMS), o coeficiente de correlação de Spearman (SP) e o erro médio absoluto (MAE), todas estas medidas aparecem definidas no Glossário. A comparação entre modelos é feita tanto ao nível geral, englobando todos os objectos do conjunto de treino (154 sais), como ao nível específico de cada tipo de anião, estando presente em cada caso o número de objectos que possuem determinado tipo de anião (n sais), neste caso, o brometo (Br^-), o cloreto (Cl^-), tetrafluoroborato (BF_4^-), hexafluorofosfato (PF_6^-), trifluorometanosulfonato (CF_3SO_3^-) e bis(trifluorometanosulfonil)imidato ($\text{N}(\text{SO}_2\text{CF}_3)_2^-$). Focando a atenção exclusivamente nos resultados gerais obtidos para cada modelo é possível observar que o modelo baseado no algoritmo da *Random Forest* permite um melhor ajuste dos pontos de fusão previstos aos pontos de fusão experimentais (**Tabela II.10**).

Ao nível da validação cruzada, relativa ao mesmo conjunto de treino, pode-se verificar que há um decaimento generalizado da correspondência entre pontos de fusão experimentais e pontos de fusão previstos por todos os modelos. Ainda assim o modelo que apresenta melhor capacidade de previsão para novos casos é o modelo *Random Forest* ($R^2 = 0,670$ e $\text{RMS} = 42,508$ °C) seguido de perto pelo modelo *Support Vector Machines* ($R^2 = 0,664$ e $\text{RMS} = 42,895$ °C), no sentido inverso, o modelo Árvore de Decisão é claramente o modelo que apresenta menor capacidade de previsão para novos sais. Olhando para cada tipo de anião separadamente é possível observar que, ao nível da análise simples do ajuste de modelo aos objectos do conjunto de treino, o modelo *Random Forest* permite um melhor ajuste considerando qualquer tipo de anião abordado neste estudo. Relativamente ao estudo de validação cruzada (5x) o modelo de *Random Forest* só supera claramente os outros modelos se estiverem em causa sais que contenham como anião cloreto (Cl^- - $R^2 = 0,420$ e $\text{RMS} = 43,062$ °C) ou tetrafluoroborato (BF_4^- - $R^2 = 0,692$ e $\text{RMS} = 14,978$ °C). Relativamente aos aniões trifluorometanosulfonato (CF_3SO_3^-) e hexafluorofosfato (PF_6^-) o modelo *Support Vector Machines* supera claramente os outros modelos ($R^2 = 0,395$ e $\text{RMS} = 42,602$ °C para o primeiro caso e $R^2 = 0,607$ e $\text{RMS} = 25,425$ °C para o segundo caso). Relativamente ao anião brometo (Br^-), o anião mais representado no conjunto de treino, o modelo SVM também permite uma melhor *performance* ($R^2 = 0,373$ e $\text{RMS} = 53,450$ °C). Relativamente aos sais contendo o anião bis(trifluorometanosulfonil)imidato ($\text{N}(\text{SO}_2\text{CF}_3)_2^-$) não há uma clara tendência para atribuir um melhor modelo

Relativamente ao estudo das características estruturais mais influentes na modelação do ponto de fusão de sais de imidazólio, não é possível uma interpretação directa pelo modelo Árvore de Decisão, já que esta apresenta uma capacidade de previsão muito reduzida para novos casos, podendo-se concluir que as regras extraídas por este modelo não são generalizáveis principalmente a objectos fora do conjunto usado para construir o modelo. Ainda assim, e porque o modelo globalmente com melhor *performance* de previsão (*Random Forest*) possui uma medida de

importância das variáveis estruturais usadas na construção desse modelo, é possível apresentar aqui quais as características estruturais mais relevantes na definição do ponto de fusão:

Tabela II.11: Os 10 descritores moleculares mais importantes na construção de modelo *Random Forest*. Para cada descritor aparece representada medida de importância %IncMSE.

Posição	Descritor	%IncMSE
1	RDF- APsH - pares de átomos a distar perto 2,8 Å	14,279
2	RDF- APsH - pares de átomos a distar perto 2,9 Å	11,797
3	QMIN-Catião	11,447
4	B1Mca	11,314
5	QMAX-Catião	11,139
6	QMAX-Anião	11,077
7	B1Mc	10,839
8	Abs (MMCatão - MMAnião)	10,810
9	RDF- APsH - pares de átomos a distar perto 2,5 Å	10,690
10	RDF- APsH - pares de átomos a distar perto 1,6 Å	10,598

Através da análise à **Tabela II.11** pode-se concluir que os descritores mais importantes na definição do ponto de fusão de sais de imidazólio são dois descritores RDF considerando a abundância de pares de átomos a distar entre si distâncias perto de duas ligações (2,8 e 2,9 Å) sem contabilizar átomos de hidrogénio, estes descritores são pesados pelas polarizabilidades atómicas de cada um dos átomos intervenientes. Estes dois descritores revelam que características estruturais como o tamanho da molécula, número de ramificações e presença de átomos polarizáveis, directamente proporcionais ao valor destes dois descritores, vão influenciar significativamente o ponto de fusão de sais de imidazólio. É de relembrar o que foi mencionado na Introdução Teórica acerca da tendência para os sais de imidazólio diminuírem o ponto de fusão com o aumento do tamanho das cadeias até as interacções de Van der Waals se tornarem preponderantes, aumentando então o ponto de fusão. Por outro lado, o aumento da ramificação provoca, em geral, um aumento do ponto de fusão pela perda de graus de liberdade conformacionais, diminuindo a entropia de fusão. Também a presença de átomos polarizáveis faz aumentar o ponto de fusão pelo incremento das forças de Van der Waals. Em 3º, 5º e 6º lugares em ordem de importância surgem a carga mínima do catião, a carga máxima do catião, e a carga máxima do anião respectivamente, que podem ser relacionados com zonas que têm maior probabilidade de estabelecer interacções Coulombicas entre o anião e o catião, influenciando o ponto de fusão. Em 4º e 7º lugares aparecem descritores que assinalam a presença de anião brometo, o que estará relacionado com a abundância de sais com elevado ponto de fusão na família dos brometos. Em 8º lugar está o valor absoluto da diferença entre a massa molar do catião e anião o que poderá ser explicado pelo facto de a diferença de tamanho entre catião e anião contribuir positivamente para a entropia de fusão diminuindo o ponto de fusão. Em 9º e em 10º lugar surgem os descritores RDF considerando a abundância de pares de átomos a distar perto de 2,5 Å (2

ligações de distância) e 1,6 Å (1 ligação de distância) respectivamente, ambos os descritores pesados pelas polarizabilidades atômicas de cada átomo envolvido. Neste caso a análise realizada aos descritores situados em 1º e 2º lugar aplica-se aos descritores situados em 9º e 10º lugar.

Selecionado o modelo que apresenta a melhor performance em termos de capacidade de previsão para novos sais (RF – **Tabela II.10**), procurou inferir-se se na realidade era uma boa opção a construção de um modelo global que incluísse diversos aniões (6 aniões diferentes neste caso) ou se pelo contrário, o recurso a modelos individuais para cada anião seria preferível. Desta forma foi feita comparação ao nível da capacidade de previsão, com recurso a um conjunto independente de teste de 79 sais, entre modelo RF global e modelos RF individuais. A comparação foi feita tanto ao nível de subconjuntos específicos, contendo um determinado tipo de anião, como ao nível global incluindo o conjunto completo de 79 sais (**Tabela II.12**). À semelhança do modelo RF global, cada modelo individual está baseado no algoritmo das *Random Forest* e foi construído com 1000 árvores e parâmetro $mtry = 25$, sendo utilizados nestes os mesmos 210 descritores referentes ao catião que foram usados no modelo global. Como o anião era uma constante em cada modelo individual, não foram usados descritores referentes ao anião.

Tabela II.12: Comparação, ao nível de conjunto de treino e teste, entre modelo RF global e modelos RF individuais para cada tipo de anião.

		Conjunto de Treino		Conjunto de Teste	
		RF global	RF individuais	RF global	RF individuais
Geral	R ²	0,985	0,988	0,747	0,733
	RMS (°C)	10,287	9,258	35,080	35,911
	SP	0,991	0,994	0,866	0,862
	MAE (°C)	6,831	6,426	28,330	28,288
	n sais	154		79	
Br ⁻	R ²	0,986	0,989	0,549	0,498
	RMS (°C)	11,872	10,620	39,240	40,918
	SP	0,994	0,995	0,729	0,727
	MAE (°C)	8,905	8,370	33,697	33,901
	n sais	62		49	
Cl ⁻	R ²	0,962	0,988	0,397	0,690
	RMS (°C)	12,544	11,543	32,466	28,544
	SP	0,977	0,974	0,821	0,786
	MAE (°C)	7,621	7,970	24,922	24,311
	n sais	20		7	
BF ₄ ⁻	R ²	0,986	0,987	0,268	0,518
	RMS (°C)	3,879	3,817	17,407	14,132
	SP	0,996	0,994	0,486	0,829
	MAE (°C)	2,776	2,784	12,473	11,695
	n sais	19		6	
PF ₆ ⁻	R ²	0,978	0,977	0,520	0,612
	RMS (°C)	7,533	6,684	20,151	21,628
	SP	0,985	0,992	0,783	0,867
	MAE (°C)	3,439	3,264	17,098	14,881
	n sais	26		9	
CF ₃ SO ₃ ⁻	R ²	0,984	0,977	0,703	0,466
	RMS (°C)	11,816	11,658	49,144	47,818
	SP	0,973	0,929	-1	-0,5
	MAE (°C)	9,937	9,805	40,159	39,614
	n sais	13		3	
N(SO ₂ CF ₃) ₂ ⁻	R ²	0,958	0,990	0,638	0,653
	RMS (°C)	7,721	4,902	18,280	19,327
	SP	0,901	0,980	0,700	0,300
	MAE (°C)	5,435	3,287	12,648	16,086
	n sais	14		5	

Tanto ao nível geral como para cada tipo de anião de cada conjunto de dados aparecem discriminadas medidas de qualidade dos modelos: R² – Quadrado do coeficiente de correlação de Pearson. RMS - Raiz do erro médio quadrático . SP – Coeficiente de Spearman. MAE – Erro médio absoluto.

Através da análise à **Tabela II.12** é possível observar que ao nível do conjunto de treino, tanto modelo RF global como os modelos RF individuais têm um excelente desempenho tanto ao

nível geral como ao nível individual para cada tipo de anião específico. Relativamente à capacidade de previsão para novos sais, pode observar-se que globalmente o modelo RF global tem capacidade de previsão ligeiramente superior que a dos modelos individuais reunidos.

Relativamente à capacidade de previsão para novos sais, esta foi avaliada através de um conjunto independente de teste de 79 sais, assim, pode-se observar através da **Tabela II.12** que ao nível geral o modelo RF global tem maior capacidade de previsão que os modelos individuais reunidos; aparece na **Figura II.17** a comparação entre os pontos de fusão experimentais e os pontos de fusão previstos pelo modelo RF global.

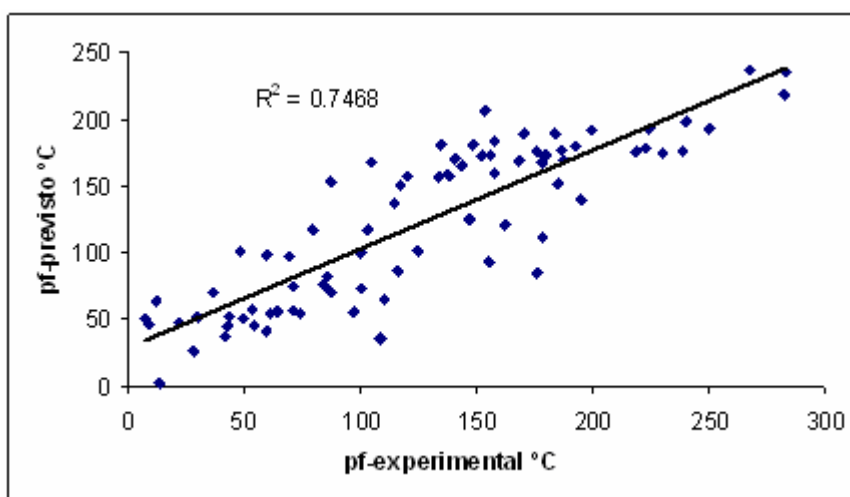


Figura II.17: Conjunto de teste de 79 sais – ponto de fusão experimental vs ponto de fusão previsto por modelo RF global e respectivo quadrado do coeficiente de correlação de Pearson R^2 .

Nesta análise há que ter em conta que tanto ao nível do conjunto de treino como ao nível do conjunto de teste o anião brometo está claramente mais representado e o modelo RF global permite melhor desempenho que o modelo RF individual relativamente ao anião brometo (Br^-). Uma tendência semelhante foi observada para o anião bis(trifluorometanosulfonil)imidato ($\text{N}(\text{SO}_2\text{CF}_3)_2^-$) enquanto que para os outros aniões o respectivo modelo individual permite melhor desempenho que o modelo global. Estes resultados podem ser afectados pela dimensão reduzida do conjunto de teste para várias famílias de anião, mas apontam para a ausência de vantagem significativa na utilização de modelos individuais.

II-3.2 - Estudo B: Classificação de Sais de Imidazólio segundo Ponto de Fusão/Estado físico

Relativamente à possibilidade de modelar o ponto de fusão de sais de imidazólio abordada na secção anterior há algumas considerações que é necessário fazer. Em primeiro lugar estando a ser estudado directamente o valor numérico do ponto de fusão é necessário possuir dados experimentais

relativos a essa propriedade. Esses dados experimentais foram retirados directamente de artigos científicos em que o ponto de fusão servia para caracterizar determinado sólido (sal de imidazólio) havendo escassos exemplos de compostos líquidos à temperatura ambiente e que tivessem ponto de fusão medido. Esta clara limitação na informação não permitiu no estudo anterior a construção de um modelo vocacionado para identificar sais de imidazólio líquidos à temperatura ambiente. Procurando resolver esta situação e obter um modelo prático que identificasse sais de imidazólio líquidos à temperatura ambiente, em vez de construir uma *regressão* para o ponto de fusão, construiu-se um modelo de classificação em termos de gamas de pontos de fusão/estado físico. Foram consideradas 3 classes: classe “a” – compostos líquidos à temperatura ambiente (ponto de fusão igual ou inferior a 25°C), classe “b” – compostos com ponto de fusão compreendido entre 25°C e 100°C, e classe “c” – compostos com ponto de fusão superior a 100°C. Neste contexto é possível incluir no estudo compostos que se sabe serem líquidos mas que não tenham um ponto de fusão associado. Este estudo foi realizado seguindo uma abordagem complementar em Químio-informática e Síntese Orgânica, sendo construídos e otimizados modelos de classificação de sais de imidazólio, que foram triados por validação cruzada (*c.v.*). O melhor modelo foi validado com conjunto independente de compostos, sendo posteriormente usado na identificação e síntese de novos líquidos iónicos.

Tabela II.13: Principais características dos modelos construídos.

Método de aprendizagem automático	Número de descritores	Tipo de descritores	Normalização dos descritores	Parâmetros associados ao método	Informação fornecida pelo modelo
<i>Random Forest</i>	281	Catião e Anião	Não	1000 árvores, <i>mtry</i> = 25	1) previsão 2) probabilidade de previsão 3) proximidade entre compostos 4) importância dos descritores
Árvore de Decisão	10	Catião e Anião	Não	referência	1) previsão 2) probabilidade de previsão 3) esquema simplificado do efeito de descritores relevantes
<i>Support Vector Machines</i>	50	Catião e Anião	Interna (normalização Z)	C = 8	1) previsão

Na **Tabela II.13** estão indicadas as características dos vários modelos construídos. Os modelos *Random Forest* são construídos partindo de um número alargado de descritores (281 neste caso), não sendo necessária uma selecção prévia de descritores. Também as Árvores de Decisão operam bem a partir de um conjunto alargado de descritores, seleccionando de uma forma explícita, simples e interpretável os descritores para a sua construção, que em geral são escassos (10 descritores neste caso). Enquanto que, tanto *Random Forest* como Árvores de Decisão operam sobre os

descritores moleculares de uma forma sequencial no estabelecimento da relação entre descritores (que representam a estrutura de determinado composto) e a propriedade/classe que se pretende modelar/prever, o método *Support Vector Machines* é sensível ao número, qualidade e inter-correlação entre os descritores visto operarem com todos os descritores em simultâneo, tornando conveniente uma selecção prévia de descritores. Relativamente ao tipo de descritores utilizados e tendo em conta o objecto de estudo nestes modelos, sais, representados por catião e anião, foi necessário usar descritores que representassem o catião e descritores que representassem o anião. Relativamente à normalização de descritores esta foi necessária em métodos que operam sobre todos os descritores em simultâneo como é o caso de *Support Vector Machines*, esta é uma forma de assegurar que todos os tipos de descritores estão compreendidos numa gama comparável de valores não havendo tipos de descritores que *a priori* tenham mais importância que outros tipos de descritores. Relativamente à informação fornecida por cada modelo é possível observar que o modelo *Random Forest* oferece mais informação complementar para além da previsão: uma medida de probabilidade das previsões indexada à proporção de árvores da *Random Forest* que vão atribuir determinada previsão, uma medida de proximidade entre as moléculas obtida através da proporção de nós terminais em que dois objectos coincidem, e uma medida de importância dos descritores moleculares usados para construir o modelo. Na mesma linha o modelo Árvore de Decisão fornece uma medida de probabilidade da previsão associada à preponderância de determinada classe num nó terminal particular, assim como um esquema interpretável do efeito dos descritores moleculares na modelação da propriedade de interesse, neste caso o estado físico/ponto de fusão. Como vai ser visto mais adiante estes modelos vão ser usados de uma forma complementar. Os modelos descritos na **Tabela II.13** foram comparados entre si, tanto na capacidade de se ajustarem às classes experimentais dos objectos do conjunto de treino, como na capacidade de previsão através de validação cruzada dividindo o conjunto de treino em 5 subconjuntos. Os resultados obtidos aparecem condensados na **Tabela II.14**.

Tabela II.14: Comparação, entre modelos RF, SVM e AD aplicados ao conjunto de treino com e sem validação cruzada 5x.

			Conjunto de treino			Validação cruzada		
						(5x)		
			RF	SVM	AD	RF	SVM	AD
Geral	Concordância		1	0,984	0,881	0,764	0,844	0,722
	k		1	0,976	0,820	0,643	0,765	0,582
	λ_B		1	0,975	0,810	0,624	0,752	0,558
	Predictibilidade	a	1	0,964	0,892	0,770	0,867	0,757
		b	1	0,992	0,917	0,675	0,792	0,616
		c	1	1	0,846	0,835	0,861	0,775
	Sensibilidade	a	1	0,994	0,914	0,827	0,883	0,790
		b	1	0,955	0,739	0,619	0,709	0,575
		c	1	1	0,979	0,829	0,929	0,821
	n objects		218					
brometo	Concordância		1	1	0,921	0,817	0,905	0,825
	N objects		63					
cloreto	Concordância		1	1	0,857	0,857	0,833	0,571
	N objects		21					
tetrafluoroborato	Concordância		1	0,976	0,927	0,854	0,878	0,841
	N objects		41					
hexafluorofosfato	Concordância		1	0,965	0,930	0,674	0,849	0,628
	N objects		43					
trifluorometanosulfonato	Concordância		1	1	0,625	0,250	0,500	0,313
	N objects		16					
bis(trifluorometanosulfonil)imidato	Concordância		1	1	0,824	0,853	0,853	0,794
	N objects		34					

Ao nível geral aparecem descritas medidas de qualidade dos modelos: concordância, k, λ_B , predictibilidade e sensibilidade. Ao nível de cada ânion aparece discriminado o respectivo valor de Concordância.

Segundo a Tabela II.14 é possível observar dois tipos diferentes de estudos, um realizado através da análise do conjunto de treino de 218 sais o outro realizado com o mesmo conjunto de sais mas através de validação cruzada considerando 5 subconjuntos do conjunto de treino (ver Validação Cruzada no glossário), este último estudo permite aferir a capacidade do modelo para prever ponto de fusão/estado físico para novos sais não utilizados no treino do modelo. Analisando as medidas estatísticas associadas ao conjunto de 218 sais como um todo é possível ver que os modelos baseados em *Random Forest* e *Support Vector Machines* se ajustam quase perfeitamente às classes de ponto de fusão/estado físico do conjunto de treino sendo obtidas concordâncias perto de 1 como seria de esperar. A validação cruzada revela que o modelo baseado em *Support Vector Machines* permitiu a

obtenção de concordância mais elevada, 0,844, um valor que indica não haver sobreajustamento significativo. Analisando este modelo mais pormenorizadamente (estudo de validação cruzada) é possível observar que as SVM permitem a obtenção de valores elevados relativamente a outras medidas estatísticas descontando o efeito de previsão aleatória ($k = 0,765$) ou o efeito de distribuição assimétrica de classes ($\lambda_B = 0,752$) ou até mesmo a sensibilidade (0,709-0,929) e predictibilidade (0,792-0,867). Olhando especificamente para cada família de aniões presente neste conjunto é possível observar que para a generalidade dos aniões o modelo baseado em *Support Vector Machines* permite a obtenção de concordâncias mais elevadas, com excepção para o anião cloreto e bis(trifluorometanosulfonyl)imidato, para os quais as RF obtiveram concordâncias iguais ou ligeiramente superiores. Pode ainda constatar-se que todos os modelos construídos têm fraca capacidade de previsão para sais de trifluormetanosulfonato (concordância = 0,250 – 0,500). Este facto poderia ser explicado pelo número reduzido de sais com este anião presentes no conjunto de treino (16 sais). Mas os sais contendo anião cloreto também são escassos (21 sais) e as concordâncias são substancialmente mais elevadas (0,571-0,857), logo esta não é a única razão nem porventura a mais significativa. A análise à **Figura II.18** que apresenta uma distribuição de sais do conjunto de treino por tipo de anião presente e classe de ponto de fusão/estado físico revela que para a generalidade dos aniões não há uma distribuição equitativa de sais pelas classes a, b e c; este facto mostra a tendência que os sais que contêm determinado anião têm para exibir determinadas gamas de ponto de fusão/estado físico. Por exemplo, os sais contendo anião brometo usados neste estudo têm clara tendência para fundir a temperaturas superiores a 100°C (classe c). Nesse aspecto o anião que exibe uma distribuição mais equitativa pelas 3 classes é o trifluormetanosulfonato, o que torna a previsão mais difícil.

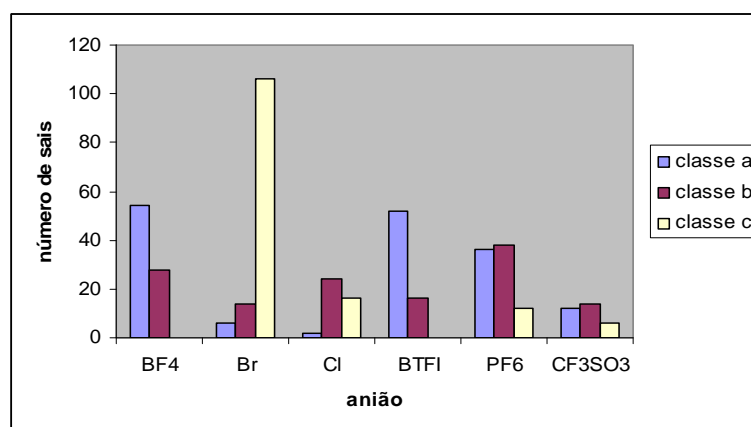


Figura II.18: Distribuição de sais por anião presente e classe ponto de fusão/estado físico. BF4 – tetrafluoroborato, Br – brometo, Cl – cloreto, BTFI – bis(trifluorometanosulfonyl)imidato, PF6 – hexafluorofosfato e CF3SO3 trifluormetanosulfonato.

Claramente o melhor modelo, ao nível de capacidade de previsão para novos casos, é aquele que se baseia em *Support Vector Machines*, este modelo foi construído com os 50 descritores moleculares os mais importantes seleccionados pelo modelo baseado em *Random Forest*. Estão compreendidos 38 descritores a representar o catião onde estão incluídos descritores RDF pesados pela polarizabilidade representando a abundância de pares de átomos distando entre si valores compreendidos entre 1,5 e 12,3 Å. Tal como aparece descrito na Introdução Teórica, o ponto de fusão depende em grande medida do tamanho e ramificação das cadeias substituintes nos sais de imidazólio (**Figura II.19**), estando codificadas estas características nos descritores RDF ainda que não de uma forma explícita,

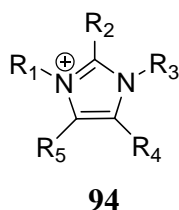


Figura II.19: Catião de imidazólio e os seus 5 grupos substituintes.

Seguindo o mesmo raciocínio foram usados outros descritores de distribuição de tamanhos das cadeias. Ainda relativamente ao catião foram seleccionados 4 descritores que codificam a superfície molecular acessível a esfera de solvente, esses descritores representam a abundância de pares de pontos da superfície do catião a distâncias entre si compreendidas entre 0 - 1 Å, 1 - 2 Å, 2 - 3 Å e 5 - 6 Å, pesados pela electronegatividade de orbitais σ . Foi ainda usado o peso molecular, a presença de hidrogénios aceitadores e a carga máxima. Por um lado o peso molecular é uma medida indirecta mas muito correlacionada com o tamanho da molécula, e como foi visto anteriormente, este influencia, em grande medida, o ponto de fusão. Por outro lado o número de hidrogénios aceitadores é uma indicação de pontos no catião com capacidade para interagir com o anião. Finalmente a carga máxima pode relacionar-se com pontos da molécula capazes de interações Coulombicas. Relativamente ao anião foram seleccionados 12 descritores estando incluídos a) a carga máxima e a carga mínima representando pontos com propensão para interagir com o catião de forma atractiva e/ou repulsiva, b) um descritor que codifica o tamanho do anião, c) dois descritores que representam a diferença de peso molecular entre catião e anião e d) descritores binários referentes à ausência ou presença de determinado anião, pesados pelo peso molecular do catião e pela diferença entre peso molecular de catião e anião.

Ainda que o modelo baseado em *Support Vector Machines* permita a obtenção de melhores resultados ao nível da capacidade de previsão este é um modelo de difícil interpretação e a relação entre descritores moleculares e propriedade físico-química de interesse é vista na prática como uma caixa negra. Por outro lado, o modelo baseado no algoritmo da Árvore de Decisão ainda que seja mais

modesto na *performance* obtida em validação cruzada é facilmente interpretável. A Árvore de Decisão obtida está na **Figura II.20**.

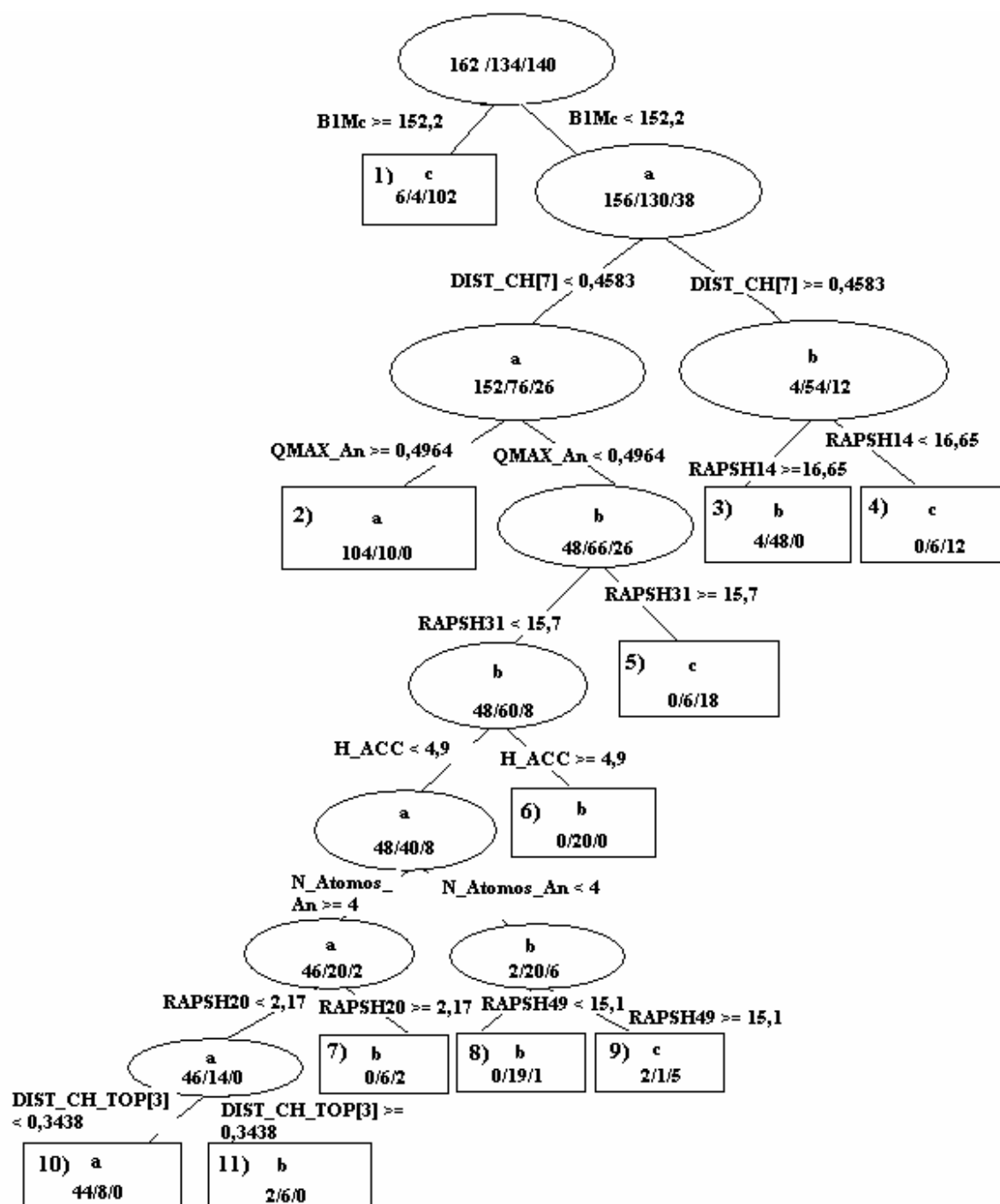


Figura II.20: Árvore de Decisão construída para previsão de classe ponto de fusão/estado físico de sais de imidazólio.

Na Árvore da **Figura II.20** pode-se observar uma primeira partição dos objectos obtida através do descritor aniônico B1Mc, este descritor representa o produto da presença (B1=1) ou ausência (B1=0) de anião brometo pelo peso molecular do catião do sal em causa. Na partição se o produto B1.Mc for superior ou igual a 152,2 significa que uma grande maioria de compostos com esta

característica vão ser de classe c, nó terminal **1**), ou seja vão ter ponto de fusão superior a 100°C. Para que $B1.Mc \geq 152,2$ o sal tem que possuir brometo como anião (caso contrário o produto seria 0) e o peso molecular do catião tem que ser superior ou igual a 152,2, ou seja o catião tem que ser pouco maior que o núcleo de imidazólio, o que acontece com grande parte dos brometos do conjunto de treino. Isto confirma que o facto de os sais de brometo serem maioritariamente da classe c e de uma grande parte dos sais de classe c serem brometos simplifica consideravelmente o problema. A seguinte partição é obtida através do descritor DIST_CH[7], que representa a abundância de combinações de átomos do catião a distar entre si 7 ligações (dividido pelo tamanho máximo de cadeia considerado neste tipo de descritores, ou seja 12 ligações). Este descritor assinala que acima de 0,4583 há uma tendência para os objectos serem de classe b (ponto de fusão entre 25 e 100°C), e abaixo desse valor a tendência é para os objectos serem de classe a.

De referir o comportamento em U observado por Holbrey e Seddon⁷⁰ observado para os *N*-alquil-*N*-metilimidazólios com o aumento do tamanho da cadeia alquílica, ou seja a partir de determinadas dimensões das cadeias o ponto de fusão começa a aumentar. Relativamente ao conjunto catalogado com classe a, este sofre partição em dois subconjuntos mais pequenos através do descritor QMAX_An (carga máxima do anião – mais carácter de catião). Desta forma sais cujo anião tenha carga máxima superior a 0,4964 vão ser de classe a, nó terminal **2**), ou seja têm ponto de fusão inferior a 25°C. Aniões nestas condições podem conduzir a interacções repulsivas com o catião conduzindo a um decréscimo do ponto de fusão, nestas condições estão o bis(trifluorometanosulfonyl)imidato e o tetrafluoroborato presentes em sais cujo catião não tenha grandes dimensões (DIST_CH[7] < 0,4583. Os nós terminais **3**) e **4**) que correspondem às classes b e c respectivamente são obtidos através do descritor RDF RAPSH14 que representa a abundância de pares de átomos do catião a distar entre si aproximadamente 1,4 Å, este descritor é de difícil interpretação já que há vários motivos possíveis para o incremento nesse descritor, tal como a presença de átomos facilmente polarizáveis, presença de ramificações ou simplesmente grandes dimensões da molécula, de acordo com **Equação II.94** do Capítulo I. De qualquer das formas para catiões com dimensões consideráveis (DIST_CH[7] ≥ 0,4583) se o descritor RAPSH4 for superior ou igual a 16,65 vai corresponder predominantemente a compostos de ponto de fusão intermédio (classe b) **3**), caso contrário há tendência para os compostos serem de classe c **4**). Nas partições seguintes o número de compostos envolvidos em cada nó terminal é baixo, assim sendo as regras obtidas poderão não ser generalizáveis.

O modelo que permitiu a obtenção de melhores resultados (modelo SVM) foi analisado mais detalhadamente através de validação externa usando um conjunto independente de 109 sais de imidazólio (218 estruturas – 2 estruturas de ressonância por cada sal), constituído por 50 brometos, 8 cloretos, 16 sais de tetrafluoroborato, 15 de bis(trifluorometanosulfonyl)imidato, 15 de

hexafluorofosfato e 5 sais de trifluormetanosulfonato. Os resultados obtidos aparecem seguidamente representados por uma tabela de contingência:

Tabela II.15: Tabela de contingência obtida por modelo SVM para conjunto de teste de 109 sais.

		Classe experimental			Total - linha
		a	b	c	
Classe prevista	a	28,5	7	2,5	38
	b	3,5	15	3	21,5
	c	3	3	43,5	49,5
	Total - coluna	35	25	49	109

Neste caso podem-se observar alguns números não inteiros porque cada sal está representado por duas estruturas de ressonância (ver secção de Metodologia Computacional) e acontece que a classe de ponto de fusão/estado físico prevista pode não ser igual para as duas estruturas de ressonância de determinado sal. Uma análise mais detalhada desta tabela e das previsões obtidas especificamente para cada família de aniões resulta na seguinte tabela de resultados:

Tabela II.16: Resultados obtidos por modelo SVM para conjunto independente de teste de 109 compostos.

			Conjunto de teste SVM
Geral	Concordância		0,798
	k		0,684
	λ_B		0,633
	Pred.	a	0,750
		b	0,698
		c	0,879
	Sens.	a	0,814
		b	0,600
		c	0,888
	N objects		109
brometo	Concordância		0,930
	N objects		50
cloreto	Concordância		0,750
	N objects		8
tetrafluoroborato	Concordância		0,781
	N objects		16
hexafluorofosfato	Concordância		0,667
	N objects		15
trifluorometanosulfonato	Concordância		0
	N objects		5
bis(trifluorometanosulfonil)imidato	Concordância		0,800
	N objects		15

Ao nível geral aparecem discriminadas medidas de qualidade dos modelos: concordância, k, λ_B , pred e sens. Ao nível de cada anião aparece discriminado o respectivo valor de Concordância. Pred. – predictibilidade, Sens. – sensibilidade.

Da análise à Tabela II.16 é possível constatar que ao nível geral os resultados são semelhantes aos obtidos por validação cruzada (Tabela II.14). A concordância é ligeiramente inferior (0,798 vs 0,844), assim como os valores de k e λ_B (0,684 vs 0,765 e 0,633 vs 0,752 respectivamente). Relativamente à predictividade e sensibilidade os resultados também são bastante coincidentes com os da validação cruzada (0,698 - 0,879 e 0,600 - 0,888 respectivamente). A previsão para objectos de classe experimental “b” é mais problemática do que a de classes extremas (“a” e “c”). Olhando especificamente para cada tipo de anião, é possível observar que o conjunto dos brometos foi previsto com concordância bastante elevada, o que se deve ao facto destes sais estarem muito representados no conjunto de treino e haver uma tendência muito marcada para compostos contendo este anião apresentarem ponto de fusão acima de 100°C. No polo oposto estão os sais de trifluorometanosulfonato que tal como na validação cruzada, são previstos com uma concordância baixa, no caso do conjunto de teste a concordância é 0.

Apesar de não ter sido seleccionado como modelo de referência para obtenção de previsões para novos casos, o modelo *Random Forest* possui duas vantagens que os outros modelos não possuem, pelo menos de uma forma tão apurada. O algoritmo da *Random Forest* providencia uma indicação da fiabilidade das previsões, através da probabilidade de pertença a determinada classe, baseada no número de votos das árvores individuais que pertencem ao *ensemble* (“floresta”). Como o número de árvores foi elevado (1000 árvores) pode considerar-se mais significativo o resultado da probabilidade. De forma a demonstrar a correlação entre a probabilidade e o grau de acerto do modelo, calculou-se a concordância associada a determinados patamares mínimos de probabilidade usados para seleccionar objectos do conjunto de teste a incluir no estudo:

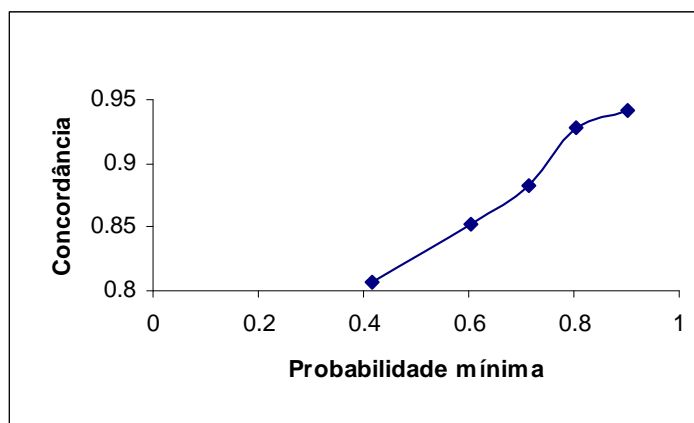


Figura II.21: Relação entre o patamar de probabilidade mínima para considerar objectos e a concordância para o conjunto de teste.

Através da **Figura II.21** pode observar-se uma elevada correlação entre o patamar de probabilidade mínima e respectiva concordância associada a determinado conjunto de objectos. Esta tendência é reprodutível no estudo de validação cruzada com 5 subconjuntos do conjunto de treino de 218 sais. O algoritmo da *Random Forest* possui outra medida bastante útil que é a proximidade entre dois objectos. Esta medida é obtida através do número de vezes que dois objectos coincidem no mesmo nó terminal em relação a todas as árvores usadas na construção do modelo RF. Neste caso para cada objecto do conjunto de teste recolheram-se medidas de proximidade em relação a objectos do conjunto de treino. Para cada objecto do conjunto de teste procurou-se a proximidade máxima e procurou observar-se qual a relação entre o patamar mínimo e a concordância:

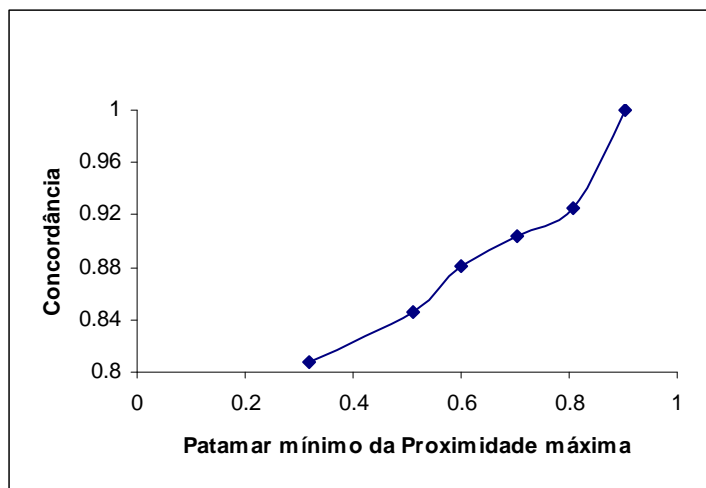


Figura II.22: Conjunto de teste - Relação entre o patamar mínimo para considerar objectos, relativamente ao valor da proximidade máxima com objectos do conjunto de treino, e a concordância.

A correlação positiva observada na **Figura II.22** é reprodutível no estudo de validação cruzada (5x) para conjunto de treino de 218 sais. Nesse caso são avaliadas as proximidades entre subconjunto deixado de fora e os restantes objectos do conjunto de treino. Assim pode concluir-se que a previsão obtida por modelo *Random Forest* é fiável para um sal que tenha probabilidade de previsão ou proximidade, em relação a objectos do conjunto de treino, elevadas.

Um dos objectivos deste estudo consistia na síntese de novos sais de imidazólio que fossem líquidos iónicos, a escolha das estruturas alvo a sintetizar baseou-se nas previsões obtidas pelos três modelos estudados, para cada uma dessas estruturas. Foram escolhidas duas estruturas alvo com potencial para serem líquidos iónicos e que foram sintetizadas:

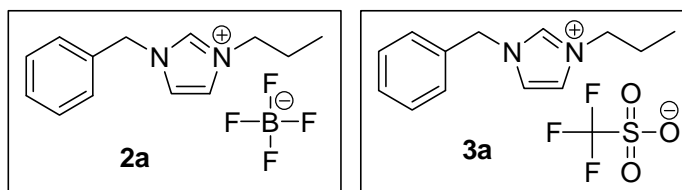
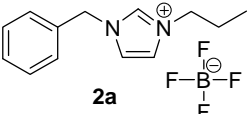
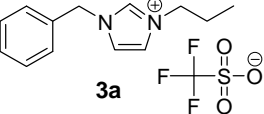


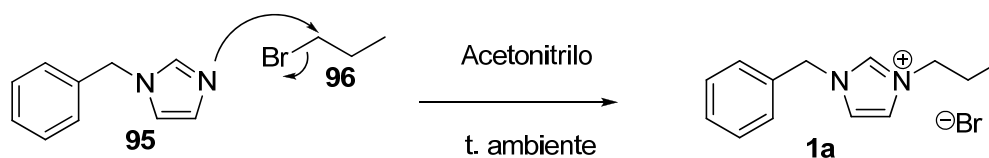
Figura II.23: Duas estruturas alvo seleccionadas pelo modelos para serem líquidos iónicos.

Relativamente às duas estruturas alvo apresentadas na **Figura II.23**, foram obtidas previsões pelos três modelos anteriormente discutidos (**Tabela II.17**), dando indicação que os dois compostos alvo têm potencial para ser líquidos iónicos:

Tabela II.17: Previsões de classe estado físico/ponto de fusão obtidas para as duas estruturas alvo.

	Classe prevista SVM	Classe prevista RF	Classe prevista AD
 <p>2a</p>	a	a	a
 <p>3a</p>	c	b	b

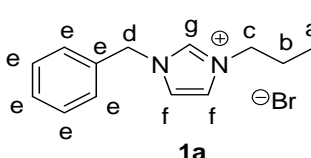
Como é possível observar pela **Figura II.23** ambas as estruturas alvo baseiam-se no mesmo catião, o 1-benzil-3-propilimidazólio diferenciando no anião. Numa primeira fase, a obtenção da estrutura de catião pretendida baseou-se na reacção de alquilação do 1-benzilimidazólio com bromopropano (os detalhes desta síntese estão descritos na Parte Experimental para a estrutura **1a**) de onde resulta o brometo de 1-benzil-3-propilimidazólio (**Esquema II.17**):



Esquema II.17: Mecanismo reaccional relativamente à obtenção de brometo de 1-benzil-3-propilimidazólio.

Este composto foi obtido como um óleo laranja viscoso com rendimento de 84,5% sendo posteriormente caracterizado (**Tabela II.18**):

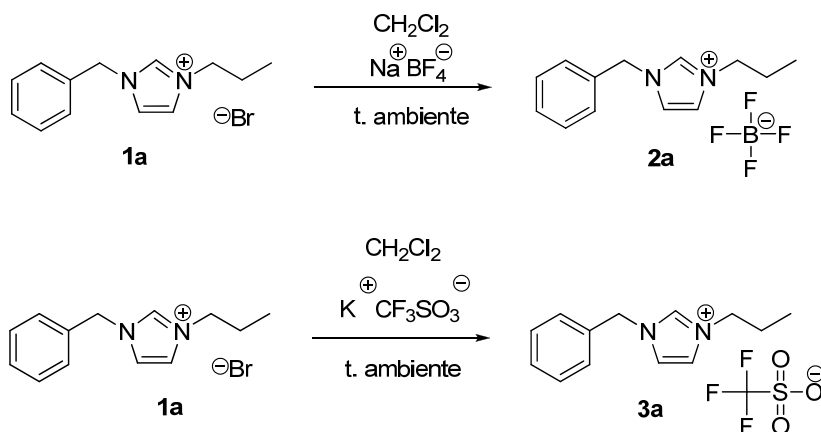
Tabela II.18: Dados reaccionais, espectroscópicos e análise elementar do brometo de 1-benzil-3-propilimidazólio (1a).

	Tempo de reacção η (%)	^1H -RMN δ (ppm) J (Hz)	^{13}C -RMN δ (ppm)	IV(filme) $\nu_{\text{máx}}$ (cm^{-1})	Análise elementar
 1a	24 horas 84,5%	0,94 (3H, t, $J=7,6$) Ha 1,93 (2H, sx, $J_1=7,2$, $J_2=7,6$) Hb 4,25 (2H, t, $J=7,2$) Hc 5,59 (2H, s) Hd 7,48-7,26 (7H, complexo) He+Hf 10,53 (1H, s) Hg	10,74 Ca 23,60 Cb 51,61 Cc 53,26 Cd 121,97 Cf 122,28 Cf 129,35 Ce 129,39 Ce 129,43 Ce 133,12 Ce 136,92 Cg	3130 (CH aromática) 3066 (CH aromática) 2968 (CH alifática) 2877 (CH alifática) 1626 (C-N) 1498 (CC aromática) 1456 (CC aromática) 1406 (CC aromática)	calculada para $\text{C}_{13}\text{H}_{17}\text{N}_2\text{Br} \cdot 0,7 \text{H}_2\text{O}$ (%): N 9,53, C 53,14, H 6,31; composição determinada (%): N 9,69, C 53,59, H 6,17.

η – rendimento (%), δ - desvio químico, J – constante de acoplamento, $\nu_{\text{máx}}$ – energia representada em número de onda.

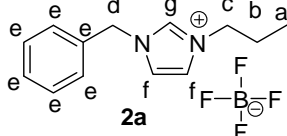
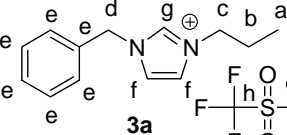
Através da Análise da Tabela II.18 é possível observar que a alquilação foi completa, ou seja, o grupo benzilo e o grupo propilo estão na proporção 1:1, já que a área dos picos Hb e Hd no espectro de ^1H -RMN são as mesmas. Toda a caracterização do brometo **1a** indica que este está puro mas não isento de água.

O brometo **1a** foi usado como precursor dos sais de tetrafluoroborato (**2a**) e trifluorometanosulfonato (**3a**):


Esquema II.18: Reacções de troca de anião para obtenção dos sais 2a e 3a obtidos a partir de brometo de 1-benzil-3-propilimidazólio 1a.

O composto **2a** foi obtido como um óleo amarelo viscoso com rendimento de troca de 80%. Por sua vez o composto **3a** foi obtido na forma de cristais brancos com rendimento de troca de 72,2%, (Pormenores relativamente à síntese destes dois compostos estão presentes na Parte Experimental, compostos **2a** e **3a**). Posteriormente estes dois compostos foram caracterizados.

Tabela II.19: Dados reaccionais, espectroscópicos e análises elementares dos sais de tetrafluoroborato (2a) e trifluorometanosulfonato (3a).

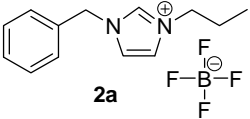
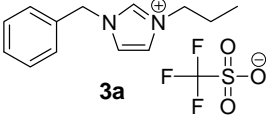
	Tempo de reacção η (%)	^1H -RMN δ (ppm) J (Hz)	^{13}C -RMN δ (ppm)	IV(filme) $\nu_{\text{máx}}$ (cm^{-1})	Análise elementar
 <p>2a</p>	24 horas 80%	0,90 (3H, t, $J=7,6$) Ha 1,89 (2H, sx, $J_1=7,2$, $J_2=7,6$) Hb 4,17 (2H, t, $J=7,2$) Hc 5,45 (2H, s) Hd 7,26-7,43 (7H, complexo) He+Hf 9,71 (1H, s) Hg	13,42 Ca 19,42 Cb 32,17 Ce 50,09 Cd 121,58 Cf 122,91 Cf 129,97 Ce 130,45 Ce 130,60 Ce 134,48 Ce 136,28 Ce 136,46 Cg	3149 (CH aromática), 3089 (CH aromática), 2970 (CH alifática), 2879 (CH alifática), 1626 (C-N) 1458 (CC aromática) 1406 (CC aromática) 1061 (BF_4^-)	calculada para $\text{C}_{13}\text{H}_{17}\text{N}_2\text{BF}_4 \cdot 0,3\text{H}_2\text{O}$ (%): N: 9,54, C: 53,20, H 6,04. composição determinada (%): N 9,35, C 53,29, H 6,62.
 <p>3a</p>	24 horas 72,2%	0,92 (3H, t, $J=7,6$) Ha 1,88 (2H, sx, $J_1=7,6$, $J_2=7,6$) Hb 4,14 (2H, t, $J=7,6$) Hc 5,35 (2H, s) Hd 7,26-7,39 (7H, complexo) He+Hf 9,27 (1H, s) Hg	10,48 Ca 23,34 Cb 51,57 Ce 53,39 Cd 122,09 Cf 122,25 Ch 122,43 Cf 128,85 Ce 129,45 Ce 129,52 Ce 132,84 Ce 136,08 Cg	3157 (CH aromática), 3095 (CH aromática), 2972 (CH alifática), 2883 (CH alifática), 1498 (CC aromática), 1456 (CC aromática), 1408 (CC aromática), 1263 (CF_3SO_3^-)	calculada para $\text{C}_{14}\text{H}_{17}\text{N}_2\text{F}_3\text{O}_3$ S (%): N: 8,00 C: 47,99, H 4,89. composição determinada (%): N 7,81, C 47,94, H 5,15.

η – rendimento (%), δ - desvio químico, J – constante de acoplamento, $\nu_{\text{máx}}$ – energia representada em número de onda.

Tendo em conta os dados reaccionais, espectroscópicos e análise elementar pode-se concluir que a estrutura **2a** foi obtida muito embora as amostras analisadas contenham água como atestam os dados da análise elementar, ainda assim pelos dados obtidos a presença de água não é tão marcada como no caso do brometo correspondente **1a**. Comparando os espectros de infravermelho de **1a** e **2a** pode-se observar neste último uma banda muito intensa a 1061 cm^{-1} que não existe no espectro de **1a**, sinal da presença de anião tetrafluoroborato, essa presença é corroborada pelos dados da análise elementar de **2a**, esses mesmos dados confirmam a pureza do composto muito embora também confirmem a presença de água. O composto **2a** foi submetido a alto vácuo (10^{-6} mbar) durante *ca* de 4h. Relativamente ao composto **3a** todos os dados espectroscópicos e análise elementar indicam a presença do composto pretendido e com elevado grau de pureza. Comparando o espectro de infravermelho do composto **1a** com o do composto **3a** pode-se observar relativamente a este último uma banda muito intensa a 1263 cm^{-1} (que não existe no espectro de IV de **1a**), evidenciando a presença de trifluorometanosulfonato, esta observação é corroborada pelos dados da análise elementar do composto **3a**. À semelhança de **1a** e **2a**, **3a** foi submetido a alto vácuo (10^{-6} mbar) durante *ca* de 4h, não evidenciando vestígios de água. Os compostos **1a**, **2a** e **3a** são novos sais estando

completamente caracterizados na Parte Experimental deste capítulo. Relativamente ao objectivo de sintetizar novos líquidos iônicos o objectivo foi cumprido (**Tabela II.20**):

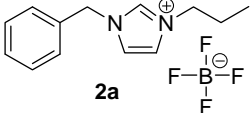
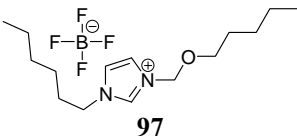
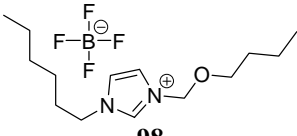
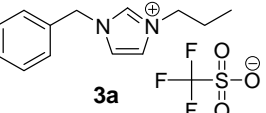
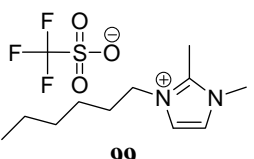
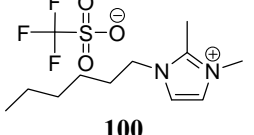
Tabela II.20: Comparação entre a classe experimental e classes previstas pelos três modelos estudados (SVM, RF e AD) relativamente ao ponto de fusão (transição vítrea) e estado físico para as duas moléculas alvo neste estudo.

	Classe prevista SVM	Classe prevista RF	Classe prevista AD	Classe experimental	p.f. (t.v.) °C	Estado físico
 2a	a	a	a	a	(-56.03)	líquido
 3a	c	b	b	b	55	sólido

Por um lado, o composto **2a** é líquido à temperatura ambiente apresentando uma *t.v.* de -56°C, ou seja é experimentalmente classe “a”, tendo sido previsto classe “a” pelos três modelos, incluído o modelo SVM. Por outro lado o composto **3a** é sólido à temperatura ambiente apresentando um ponto de fusão de 55°C sendo experimentalmente um composto de classe “b”, por sua vez dois dos três modelos construídos prevêem que o composto **3a** é de classe “b” enquanto que o outro modelo prevê que **3a** é de classe “c”. Assim, o composto **3a** é um líquido iónico já que funde a uma temperatura inferior a 100°C. Ainda relativamente às previsões obtidas para este composto, estas à partida eram pouco credíveis, dada a baixa concordância associada a este tipo de anião (trifluorometanosulfonato) para qualquer dos modelos construídos (**Tabela II.14**).

Ainda que o modelo *Random Forest* (RF) não permita a obtenção das proporções mais elevadas de acertos (**Tabela II.14**) fornece informação muito útil relativamente às previsões obtidas, tal como a probabilidade de determinadas previsões, objectos mais próximos no conjunto de treino e respectivo valor de proximidade tal como aparece na **Tabela II.21** para os dois líquidos iónicos sintetizados **2a** e **3a**. Assim, é possível ter um grau de confiança nas previsões obtidas.

Tabela II.21: Informação adicional fornecida por modelo RF.

	Classe prevista RF	Estruturas mais próximas conjunto de treino	Classe experimental estruturas mais próximas	Proximidade	Probabilidade
 2a	a	 97	a	0,432	0,685
		 98	a	0,430	
 3a	b	 99	b	0,310	0,518
		 100	b	0,306	

Para os compostos sintetizados 2a e 3a são apresentadas as estruturas mais próximas no conjunto de treino, respectivas classes experimentais e valores de proximidade tal como o valor de probabilidade.

Percorrendo a **Tabela II.21** pode-se observar que para o sal sintetizado 2a as estruturas mais próximas no conjunto de treino ainda que os aniões sejam também tetrafluoroboratos as estruturas dos catiões são significativamente diferentes em relação ao catião do sal sintetizado como indicam os valores de proximidade 0,432 e 0,430 (numa escala de 0-1). O mesmo se pode dizer em relação ao composto 3a cujas estruturas mais próximas no conjunto de treino têm também como anião o trifluorometanosulfonato ainda que as estruturas dos respectivos catiões divirjam significativamente, esse facto aparece reflectido nos baixos valores de proximidade (0,310 e 0,306). Relativamente às probabilidades das previsões obtidas estas não são excessivamente baixas tendo em conta os valores apresentados de 0,685 e 0,518 (numa escala de 0-1) e o facto deste modelo considerar três classes (a, b e c)

O brometo sintetizado **1a**, precursor de **2a** e **3a**, não foi usado na avaliação da capacidade de previsão dos modelos construídos no seguinte contexto, **1a** possui percentagem muito elevada de água, mais precisamente 9,1% em massa mesmo após período de 4 horas em que o composto esteve submetido a alto vácuo (10^{-6} mbar), assim o composto apresenta-se líquido à temperatura ambiente. Após período em que o composto foi submetido a alto vácuo apareceram pequenos cristais, muito embora em quantidade insuficiente para medir o ponto de fusão. O modelo SVM previu que o brometo **1a** ia ser de classe “c”, ou seja, vai apresentar um ponto de fusão acima de 100°C. Pela observação deste exemplo pode ilustrar-se a dificuldade em trabalhar com este tipo de sais pois as suas propriedades físico-químicas são adulteradas pela presença de determinadas impurezas como é o caso da água em sais higroscópicos.

II-3.3 - Estudo C: Modelação da Densidade de Sais de Imidazólio

Este é um estudo de regressão envolvendo sais de imidazólio, à semelhança do primeiro **II-3.1**, mas em vez do ponto de fusão foi modelada a densidade. Foram construídos três modelos distintos baseados em métodos de aprendizagem automática, sendo seleccionado o modelo com melhor desempenho em ensaio de validação cruzada ($3\times$, 3 subconjuntos do conjunto de treino). Esse modelo foi validado através de conjunto independente de sais. Serão igualmente focadas quais as variáveis estruturais mais relevantes na definição da densidade. Todos os compostos usados neste estudo são líquidos à temperatura ambiente. Os três modelos construídos inicialmente apresentam as seguintes características:

Tabela II.22: Características dos três modelos construídos para modelar a densidade de sais de imidazólio.

Método de aprendizagem automático	Número de descritores em que se baseia o modelo	Tipo de descritores	Normalização	Parâmetros associados ao método
<i>Random Forest</i>	246	Catião e Anião	Não	1000 árvores, <i>mtry</i> = 10
Árvore de Decisão	3	Catião e Anião	Não	referência
<i>Support Vector Machines</i>	60	Catião e Anião	Interna (normalização Z)	$\nu = 0,016$ $C=16$

Analisando a **Tabela II.22** é possível ver que o modelo *Random Forest* foi construído através do recurso a um número elevado de descritores (246) correspondendo tanto a características estruturais do catião como do anião. Por outro lado o modelo baseado no algoritmo da Árvore de Decisão seleccionou simplesmente 3 descritores moleculares, estando contempladas tanto características estruturais do catião como do anião. Baseado em 60 descritores moleculares, o modelo

Support Vector Machines contempla tanto características estruturais de catião como de anião, os 60 descritores seleccionados são os mais importantes segundo modelo *Random Forest*.

Na fase inicial os três modelos construídos foram comparados ao nível da capacidade de ajuste às densidades experimentais do conjunto de treino, tal como ao nível de capacidade de previsão para novos sais em ensaio de validação cruzada, estando os resultados representados na **Tabela II.23**. De referir que nos três modelos construídos estavam contemplados 6 tipos de aniões diferente, o tetrafluoroborato (BF_4^-), o hexafluorofosfato (PF_6^-), o bis(trifluorometanosulfonil)imidato ($\text{N}(\text{SO}_2\text{CF}_3)_2^-$), o trifluorometanosulfonato (CF_3SO_3^-), o trifluoroacetato (CF_3CO_2^-) e o ácido-2-hidroxipropiónico ($\text{CH}_3\text{CH}(\text{OH})\text{CO}_2^-$).

Tabela II.23: Comparação do desempenho de três modelos distintos (RF, SVM e AD) relativamente a conjunto de treino com e sem validação cruzada 3x.

		Conjunto de treino			Validação cruzada (3x)		
		RF	SVM	AD	RF	SVM	AD
Geral	R ²	0,996	1,000	0,928	0,918	0,940	0,780
	RMS (g/cm ³)	0,027	0,009	0,102	0,125	0,106	0,179
	SP	0,997	0,997	0,769	0,940	0,969	0,863
	MAE (g/cm ³)	0,016	0,008	0,075	0,074	0,045	0,111
	n sais	68			68		
BF ₄ ⁻	R ²	0,991	0,990	0,537	0,819	0,820	0,554
	RMS (g/cm ³)	0,010	0,009	0,062	0,045	0,037	0,086
	SP	0,994	0,992	-0,235	0,942	0,827	0,646
	MAE (g/cm ³)	0,008	0,008	0,063	0,037	0,022	0,059
	n sais	18			18		
PF ₆ ⁻	R ²	0,995	0,994	0,808	0,946	0,893	0,796
	RMS (g/cm ³)	0,013	0,009	0,066	0,052	0,036	0,092
	SP	1,000	0,991	0,750	0,973	0,927	0,882
	MAE (g/cm ³)	0,012	0,008	0,054	0,045	0,021	0,079
	n sais	11			11		
N(SO ₂ CF ₃) ₂ ⁻	R ²	0,994	1,000	0,893	0,874	0,886	0,721
	RMS (g/cm ³)	0,044	0,009	0,154	0,205	0,186	0,254
	SP	0,993	1,000	0,778	0,907	0,962	0,738
	MAE (g/cm ³)	0,027	0,008	0,115	0,129	0,087	0,169
	n sais	20			20		
CF ₃ SO ₃ ⁻	R ²	0,999	0,991	0,842	1,000	0,935	0,600
	RMS (g/cm ³)	0,009	0,010	0,039	0,043	0,029	0,355
	SP	1,000	1,000	0,750	1,000	1,000	0,500
	MAE (g/cm ³)	0,008	0,010	0,029	0,040	0,028	0,245
	n sais	3			3		
CF ₃ CO ₂ ⁻	R ²	0,991	0,963	0,915	0,813	0,606	0,997
	RMS (g/cm ³)	0,026	0,009	0,065	0,079	0,033	0,239
	SP	1,000	1,000	0,750	1,000	0,500	1,000
	MAE (g/cm ³)	0,026	0,008	0,051	0,072	0,025	0,171
	n sais	3			3		
CH ₃ C(OH)CO ₂ ⁻	R ²	0,983	0,958	0,541	0,938	0,514	0,723
	RMS (g/cm ³)	0,021	0,009	0,084	0,099	0,051	0,093
	SP	1,000	0,956	-0,060	0,984	0,533	0,742
	MAE (g/cm ³)	0,014	0,008	0,075	0,071	0,039	0,077
	n sais	13			13		

Tanto ao nível geral como para cada tipo de anião de cada conjunto de dados aparecem discriminadas medidas de qualidade dos modelos: R² - Quadrado do coeficiente de correlação de Pearson. RMS - Raiz do erro médio quadrático . SP - Coeficiente de Spearman. MAE - Erro médio absoluto.

Através da análise à **Tabela II.23** pode-se ver que o modelo SVM é aquele que se ajusta mais aos dados experimentais do conjunto de treino permitindo a obtenção de $R^2 = 1,000$ RMS = 0,000.

Ao nível da capacidade de previsão para novos casos dada através de ensaio de validação cruzada (3x) ficou evidente que o modelo SVM é o melhor ao nível geral permitindo $R^2 = 0,940$, RMS = 0,106 g/cm³, SP = 0,969 e MAE = 0,045 g/cm³ e também individualmente para sais de tetrafluoroborato (BF₄⁻) e bis(trifluorometanosulfonyl)imidato (N(SO₂CF₃)₂⁻). Já para sais de trifluoroacetato (CF₃CO₂⁻), 2-hidroxiopropionato (CH₃C(OH)CO₂⁻), hexafluorofosfato (PF₆⁻) e trifluorometanosulfonato (CF₃SO₃⁻) a qualidade das previsões é aproximadamente a mesma para SVM e RF. Ainda que o modelo SVM permita melhor desempenho, qualquer dos modelos construídos permitem um ajuste bastante rigoroso aos dados experimentais tal como uma boa capacidade de previsão de densidade para novos sais.

Relativamente às características estruturais que influenciam a densidade de sais de imidazólio, a AD é o modelo mais simples e baseia-se em três descritores moleculares (**Figura II.24**).

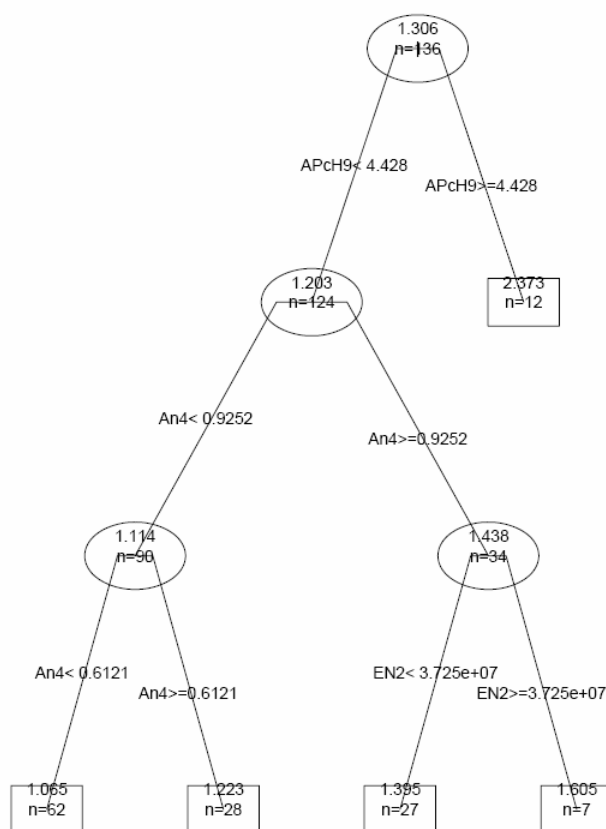


Figura II.24: Figura esquemática do modelo Árvore de Decisão usado na modelação da densidade de sais de imidazólio.

A primeira partição é obtida através do descritor APcH9, que é um descritor RDF considerando átomos de hidrogénio e que representa a abundância de pares de átomos a distar perto de 0,9 Å no catião tendo em conta a polarizabilidade dos átomos intervenientes no valor final do descritor. A partir desta partição é possível isolar um grupo de estruturas num nó terminal cuja média de densidades é 2,373 g/cm³ correspondendo a estruturas cujo descritor APcH9 é igual ou superior a 4,428. Na prática as estruturas isoladas neste grupo correspondem a sais contendo no catião elementos muito polarizáveis como é o caso do bromo ou do iodo sendo estes elementos bastante pesados, contribuindo para as elevadas densidades que estes sais possuem. Por outro lado, formou-se outro grupo nessa partição correspondendo a uma média de densidades significativamente mais baixa (1,203 g/cm³). Para uma estrutura pertencer a este grupo o descritor APcH9 tem que ser inferior a 4,428. Este grupo é repartido noutros grupos mais pequenos, num dos grupos o descritor An4 têm que ser igual ou superior a 0,9252, correspondendo a densidades em média mais elevadas (1,438 g/cm³), no outro grupo An4 tem que ser inferior a 0,9252, correspondendo a densidades em média mais baixas (1,114 g/cm³). O descritor An4 representa o quociente entre peso molecular do anião e peso molecular do catião. Objectivamente esta partição mostra a tendência de sais com tamanhos de catião e anião semelhantes para apresentarem densidades mais elevadas do que sais em que essa relação é mais desigual.

Por sua vez o grupo com média de densidades igual a 1,114 g/cm³ sofrer partição em dois grupos mais pequenos, responsável por esta partição aparece novamente o descritor An4 (quociente entre pesos moleculares de anião e catião), assim, estruturas cujo quociente An4 seja igual ou superior a 0,6121 vão ser colocadas num grupo cuja média de densidades é 1,223 g/cm³. Por sua vez, quando o valor do quociente An4 nas estruturas é inferior a 0,6121, estas vão ser colocadas num grupo cuja média de densidades é 1,065 g/cm³, novamente fica evidente a importância da simetria de tamanhos entre catião e anião para que a densidade seja mais elevada. Estes dois últimos grupos são nós terminais correspondendo a sais cujos catiões não possuem átomos pesados como é o caso do bromo ou o iodo, havendo alguma assimetria entre o tamanho de catião e anião.

Voltando ao grupo cuja média de densidades é 1,438 g/cm³, este sofre partição em dois grupos ainda mais pequenos, responsável por esta partição está o descritor EN2, representando a abundância de pares de pontos na superfície de determinado catião a distar entre si 1-2 Å, esta abundância é obtida em relação ao número total de distâncias a considerar entre pares de pontos dessa superfície, o valor final deste descritor é directamente proporcional à electronegatividade sigma associada aos pontos intervenientes nesse cálculo. Assim, estruturas cujo valor de EN2 seja inferior a $3,725 \times 10^7$ as respectivas estruturas vão ser colocadas em nó terminal cuja média de densidades é 1,395 g/cm³.

No caso de estruturas cujo valor de EN2 é igual ou superior a $3,725 \times 10^7$ estas vão ser colocadas em nó terminal cuja média de densidades é $1,605 \text{ g/cm}^3$, neste último nó estão colocados objectos ramificados e/ou que possuem elementos electronegativos como é o caso do flúor. Os objectos agrupados nestes dois últimos nós terminais não possuem elementos pesados na estrutura do catião, e o tamanho de catião e anião no sal é semelhante. De referir que o modelo AD apesar de ser bastante simples e incorporar apenas três características estruturais, consegue ajustar-se de forma bastante rigorosa às densidades dos sais do conjunto de treino, apresentando uma capacidade de previsão bastante aceitável como demonstram os resultados da validação cruzada expressos na **Tabela II.23**.

O modelo *Random Forest* parte do conjunto total de 246 descritores, sendo estes usados, à semelhança do modelo AD, de forma sequencial na construção do modelo RF final. Os 10 descritores mais importantes segundo este modelo estão na **Tabela II.24**.

Tabela II.24: Os 10 descritores mais importantes na modelação da densidade de sais de imidazólio segundo modelo Random Forest.

Posição	Descritor	%IncMSE
1	An42 - produto do descritor binário, que indica a presença de tetrafluoroborato como anião, pela razão entre peso molecular de anião e catião	8,069
2	XP7 - Potencial máximo de aceitador de hidrogénios (catião)	7,898
3	APcH16 - Descritor RDF considerando a polarizabilidade de átomos (sem contabilizar átomos de hidrogénio) que distem entre si perto de 1,6 Å (catião)	7,677
4	An3 - Valor absoluto da diferença entre peso molecular do catião e anião	7,508
5	An35 - produto do descritor binário, que indica a presença de 2-hidroxi-propionato como anião, pelo valor absoluto da diferença entre peso molecular de anião e catião	7,461
6	XP41 – DIST_Q[6] (catião)	7,453
7	An37 - produto do descritor binário, que indica a presença de hexafluorofosfato como anião, pela razão entre peso molecular de anião e catião	7,429
8	An23 - produto do descritor binário, que indica a presença de 2-hidroxi-propionato como anião, pelo valor do peso molecular de anião	7,416
9	An30 - produto do descritor binário, que indica a presença de tetrafluoroborato como anião, pelo valor absoluto da diferença entre peso molecular de anião e catião	7,370
10	EN1 - Descritor SURFACE considerando pontos a distar entre si 0-1 Å (catião)	7,245

Para cada descritor aparece representada medida de importância %IncMSE.

Na **Tabela II.24** é possível observar que cinco dos dez descritores mais importantes segundo modelo RF estão de alguma forma relacionados com a relação entre peso molecular de catião e anião,

estando colocado na primeira posição o descritor An42. Este descritor representa o produto do descritor binário, que indica a presença de tetrafluoroborato como anião, pela razão entre peso molecular de anião e catião. O anião tetrafluoroborato é um dos aniões mais representados no conjunto de treino. Na quarta posição surge o descritor An3 que representa o valor absoluto da diferença entre peso molecular do catião e anião, um claro indicativo de que a relação entre o peso molecular destes dois componentes é muito relevante para a densidade, como já havia sido observado na análise à Árvore de Decisão. Na quinta posição surge o descritor An35 que representa o produto do descritor binário, que indica a presença de 2-hidroxi-propionato como anião, pelo valor absoluto da diferença entre peso molecular de anião e catião. O anião 2-hidroxi-propionato é um dos aniões mais representados no conjunto de treino. Na sétima posição surge o descritor An37 que indica o produto do descritor binário, que indica a presença de hexafluorofosfato como anião, pela razão entre peso molecular de anião e catião. Relativamente à nona posição esta aparece representada pelo descritor An30 que representa o produto do descritor binário, que indica a presença de tetrafluoroborato como anião, pelo valor absoluto da diferença entre peso molecular de anião e catião - novamente um descritor relacionado com o anião tetrafluoroborato. Os restantes cinco descritores estão relacionados com o anião ou com o catião em exclusivo. Assim na oitava posição surge descritor An23 que representa o produto do descritor binário que indica a presença de 2-hidroxi-propionato como anião, pelo valor do peso molecular de anião. Os restantes quatro descritores referem-se em exclusivo ao catião, surgindo na segunda posição o descritor XP7 - potencial máximo de aceitador de hidrogénios (relacionado com possíveis interacções, nomeadamente por pontes de hidrogénio, com aniões ou outros catiões). Na terceira posição surge descritor APcH16 (catião) - descritor RDF considerando a polarizabilidade de átomos (sem contabilizar átomos de hidrogénio) que distem entre si perto de 1,6 Å (comprimento aproximado de uma ligação química). A polarizabilidade atómica é uma propriedade muito importante na definição da densidade pois está relacionada com as interacções inter-estruturas por um lado, e pelo facto de átomos muito polarizáveis como é o caso do bromo ou iodo (átomos pesados) provocarem um grande acréscimo da densidade. Na sexta posição surge descritor XP41 – DIST_Q[6] que representa uma medida de distribuição de cargas pela estrutura do catião.

Segundo os resultados da **Tabela II.23** o modelo que melhor se ajusta às densidades do conjunto de treino e que apresentava melhor capacidade de previsão, é o modelo baseado no algoritmo das *Support Vector Machines* (SVM), este modelo foi ainda validado recorrendo a um conjunto independente de teste de 34 sais para a obtenção de previsões, os resultados encontram-se resumidos na **Tabela II.25**.

Tabela II.25: Resultados obtidos por modelo SVM na previsão de densidade para conjunto independente de teste de 34 sais de imidazólio.

		Conjunto de teste
		SVM
Geral	R ²	0,975
	RMS (g/cm ³)	0,059
	SP	0,990
	MAE (g/cm ³)	0,030
	n sais	34
BF ₄ ⁻	R ²	0,954
	RMS (g/cm ³)	0,027
	SP	0,983
	MAE (g/cm ³)	0,022
	n sais	9
PF ₆ ⁻	R ²	0,936
	RMS (g/cm ³)	0,019
	SP	1,000
	MAE (g/cm ³)	0,016
	n sais	5
N(SO ₂ CF ₃) ₂ ⁻	R ²	0,960
	RMS (g/cm ³)	0,102
	SP	0,983
	MAE (g/cm ³)	0,055
	n sais	9
CF ₃ SO ₃ ⁻	R ²	0,966
	RMS (g/cm ³)	0,010
	SP	1,000
	MAE (g/cm ³)	0,008
	n sais	3
CF ₃ CO ₂ ⁻	R ²	-
	RMS (g/cm ³)	-
	SP	-
	MAE (g/cm ³)	-
	n sais	1
CH ₃ C(OH)CO ₂ ⁻	R ²	0,950
	RMS (g/cm ³)	0,046
	SP	0,964
	MAE (g/cm ³)	0,027
	n sais	7

Tanto ao nível geral como para cada tipo de anião aparecem descritas medidas de qualidade dos modelos: R² - Quadrado do coeficiente de correlação de Pearson. RMS - Raiz do erro médio quadrático. SP - Coeficiente de Spearman. MAE - Erro médio absoluto.

Através da análise aos resultados da **Tabela II.25** é possível observar que há uma excelente correlação ao nível geral entre valores previstos por modelo SVM e respectivos valores experimentais de densidade para conjunto de teste ($R^2 = 0,975$ – **Figura II.25**).

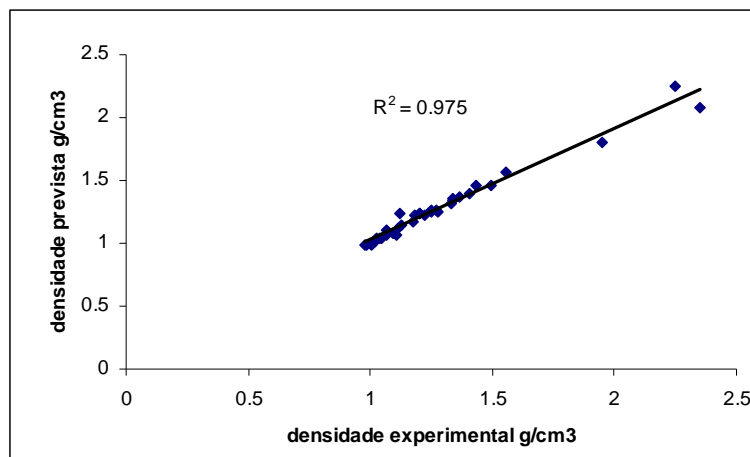


Figura II.25: Conjunto de teste – correlação entre densidades experimentais e densidades previstas por modelo SVM.

Para este conjunto de teste são obtidos desvios muito baixos ($RMS = 0,059 \text{ g/cm}^3$). Neste caso os resultados são ainda melhores do que no caso do estudo de validação cruzada ($R^2 = 0,940$ e $RMS = 0,106 \text{ g/cm}^3$ – **Tabela II.23** ao nível geral). Ao nível de cada anião os resultados também são muito bons - a correlação mais baixa foi obtida por sais contendo hexafluorofosfato (PF_6^-) como anião ($R^2 = 0,936$) e a correlação mais elevada foi obtida por sais de trifluorometanosulfonato (CF_3SO_3^-) ($R^2 = 0,966$). Os sais de bis(trifluorometanosulfonil)imidato apresentam o valor de RMS mais elevado ($0,102 \text{ g/cm}^3$) enquanto que os sais de trifluorometanosulfonato apresentam o valor de RMS mais baixo ($0,010 \text{ g/cm}^3$). Relativamente ao anião trifluoroacetato (CF_3CO_2^-), só existe um sal com este anião no conjunto de teste não sendo determinados os parâmetros de qualidade de previsão para este sal.

II-3.4 - Estudo D: Modelação do Ponto de Fusão de Sais de Guanidínio

Sais de guanidínio constituem uma nova classe de líquidos iónicos com propriedades de solubilidade e estabilidade complementares⁴² aos amplamente estudados sais de imidazólio. Para a obtenção de sais de guanidínio líquidos à temperatura ambiente ou com um ponto de fusão baixo é necessário considerar os factores estruturais relevantes. Nesse sentido foram construídos modelos baseados em redes neuronais de *Counter-Propagation*. Este tipo de modelos pode estabelecer relações não-óbvias entre factores estruturais e determinada(s) propriedade(s) de interesse (neste caso o ponto de fusão). O mapeamento de estruturas de acordo com factores estruturais permite um acesso intuitivo visual a factores estruturais que podem significativamente influenciar o ponto de fusão. Os modelos baseados em redes de *Counter-Propagation* podem assim ser bastante úteis ao químico orgânico que pretende sintetizar determinado composto com uma propriedade determinada. Assim, foram construídos dois modelos com as seguintes características:

Tabela II.26: Modelos *Counter-Propagation* base para modelação do ponto de fusão de sais de guanidínio.

Método de aprendizagem automático	Número de descritores	Tipo de descritores	Normalização de descritores	Parâmetros associados ao método
Rede Neuronal <i>Counter-Propagation</i>	92	Catião	linear	Dimensões 18x18, 92 camadas de <i>input</i> e 4 camadas de <i>output</i>
<i>Ensemble</i> de 5 Redes Neuronais de <i>Counter-Propagation</i>	92	Catião	linear	Cada uma das 5 redes tem dimensões 18x18, 92 camadas de <i>input</i> e 4 camadas de <i>output</i>

Os dois modelos referidos na **Tabela II.26** foram comparados tanto na capacidade de modelação dos pontos de fusão experimentais do conjunto de treino como na capacidade de previsão para novos sais não contemplados no conjunto de treino. Para a concretização desses objectivos foi estudado o conjunto de treino e um conjunto de teste. Ambos os conjuntos foram constituídos por sais contendo como aniões o cloreto (Cl^-), o tetrafenilborato (BPh_4^-), o brometo (Br^-) e o iodeto (I^-). Na selecção do conjunto de teste houve o cuidado de excluir qualquer objecto que fosse mapeado em alguma zona isolada do mapa da rede de *Counter-Propagation*. O modelo baseado num *ensemble* de 5 redes CPG foi construído no intuito de evitar flutuações induzidas por factores aleatórios influenciando o treino de uma única rede CPG. Os resultados estão apresentados na **Tabela II.27**.

Tabela II.27: Comparação entre modelos baseados numa rede CPG-NN e num ensemble de 5 redes CPG-NN

		Conjunto de treino		Conjunto de teste	
		CPG-NN	Ensemble 5 CPG-NN	CPG-NN	Ensemble 5 CPG-NN
Geral	R ²	0,866	0,865	0,791	0,815
	RMS (°C)	29,490	30,10	25,480	23,950
	SP	0,920	0,928	0,865	0,815
	MAE (°C)	19,960	20,430	22,010	20,330
	n sais	81		20	
Cl ⁻	R ²	0,864	0,866	0,761	0,670
	RMS (°C)	34,620	36,780	22,150	25,210
	SP	0,880	0,883	0,905	0,786
	MAE (°C)	23,130	25,900	18,650	20,370
	n sais	31		8	
BPh ₄ ⁻	R ²	0,798	0,849	0,880	0,909
	RMS (°C)	24,030	21,360	30,870	25,780
	SP	0,877	0,786	0,857	0,886
	MAE (°C)	18,870	16,330	27,580	23,160
	n sais	15		6	
Br ⁻	R ²	0,926	0,895	-	-
	RMS (°C)	11,850	13,920	-	-
	SP	0,942	0,892	-	-
	MAE (°C)	8,660	9,240	-	-
	n sais	9		2	
I ⁻	R ²	0,776	0,789	0,721	0,903
	RMS (°C)	30,510	29,640	24,230	20,300
	SP	0,808	0,828	0,400	0,800
	MAE (°C)	21,380	20,160	22,100	18,060
	n sais	26		4	

Conjuntos de treino e teste. Tanto ao nível geral como para cada tipo de anião aparecem discriminadas medidas de qualidade dos modelos: R² - Quadrado do coeficiente de correlação de Pearson. RMS - Raiz do erro médio quadrático. SP - Coeficiente de Spearman. MAE - Erro médio absoluto.

Tanto o modelo baseado numa rede CPG como o modelo baseado num *ensemble* de 5 redes CPG, reproduzem bem os valores de ponto de fusão do conjunto de treino, apresentando R² muito similares (0,866 vs 0,865 para uma rede CPG e para o *ensemble* de cinco redes CPG respectivamente). Relativamente ao conjunto de teste, o modelo baseado num *ensemble* de 5 redes CPG proporcionou R² mais elevados e RMS mais baixos do que no caso do modelo baseado numa rede (R² = 0,815 e RMS = 23,950 °C vs R² = 0,791 e RMS = 25,480 °C respectivamente).

Ainda relativamente ao conjunto de teste mas analisando cada tipo de anião individualmente é possível ver que o modelo baseado num *ensemble* de 5 redes permite correlações mais elevadas e medidas de desvios mais baixas no caso dos aniões tetrafenilborato (BPh_4^-) e iodeto (I^-) em relação ao modelo baseado numa rede neuronal CPG. No caso do anião ser cloreto (Cl^-), o modelo baseado numa rede neuronal CPG apresenta melhor desempenho. Ainda assim o *ensemble* de 5 redes permite uma maior constância nos desvios (RMS e MAE) associados a cada um dos aniões. Desta forma pode considerar-se que o modelo baseado num *ensemble* de 5 redes neuronais CPG é mais robusto e tem maior capacidade de previsão que o modelo baseado numa única rede neuronal CPG. Na **Figura II.26** representam-se graficamente os pontos de fusão experimentais vs. pontos de fusão previstos pelo modelo baseado no *ensemble* de cinco redes neuronais CPG.

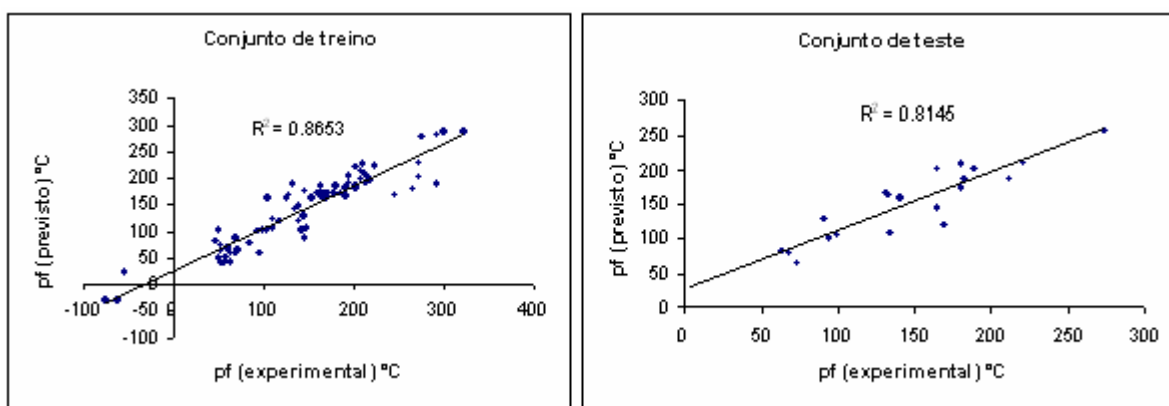


Figura II.26: Comparação entre pontos de fusão experimentais e pontos de fusão previstos por modelo baseado em *ensemble* de 5 redes CPG – Conjunto de treino e conjunto de teste. Em ambos os casos é determinado o quadrado do coeficiente de correlação de Pearson R^2 .

Posteriormente o modelo baseado num *ensemble* de cinco redes foi adicionalmente validado através de validação cruzada dividindo o conjunto de treino em cinco subconjuntos (**Tabela II.28**). Nesta validação foram descartadas as previsões correspondentes a objectos com valor de ponto de fusão superior ou inferior ao intervalo de pontos de fusão contidos no conjunto de treino dentro de determinada família de anião.

Tabela II.28: Resultados obtidos por modelo baseado em ensemble de 5 redes neuronais CPG em ensaio de validação cruzada 5x.

	Validação cruzada 5x
R ²	0,747
RMS (°C)	33,240
SP	0,793
MAE (°C)	24,699
n sais	81

R² - Quadrado do coeficiente de correlação de Pearson. RMS - Raiz do erro médio quadrático. SP - Coeficiente de Spearman. MAE - Erro médio absoluto.

Em validação cruzada o modelo baseado em *ensemble* de cinco redes CPG permitiu boas correlações ($R^2 = 0,747$ e $SP = 0,793$) tal como a obtenção de valor de RMS equivalente a 33,2°C. Este modelo foi ainda validado distribuindo de forma aleatória os valores de ponto de fusão dos objectos do conjunto de treino e treinando assim 5 redes neuronais CPG. Cada rede assim treinada foi usada para prever os pontos de fusão relativos aos objectos do conjunto de teste. No final foram obtidas correlações extremamente baixas entre pontos de fusão experimentais e pontos de fusão previstos para cada um dos 5 modelos CPG construídos: $R^2 = 0,063$; 0,2731; 0,010; 0,043 e 0,045. Este ensaio demonstra que os modelos referidos na **Tabela II.27** não assentam em factores aleatórios ocasionais.

Apesar dos modelos construídos apresentarem boa capacidade de previsão para conjunto de teste, estes assentam em 92 descritores catiónicos escolhidos manualmente. Assim sendo procurou-se inferir da possibilidade de obtenção de um melhor modelo através de selecção automática de descritores treinando um modelo com menos descritores. Para esse fim utilizou-se uma rede neuronal de *Kohonen*, a que foi submetida matriz transposta do conjunto de treino em que cada descritor (assim como o ponto de fusão) passou a ser um objecto e cada sal uma variável de *input*. Com este conjunto transposto foi treinada rede neuronal de *Kohonen* 10x10 para mapear cada um dos 92 descritores e o ponto de fusão e para posteriormente seleccionar um descritor por cada neurónio ocupado, obtendo no total um conjunto de 16 descritores. Estes 16 descritores seleccionados foram usados para representar cada um dos sais do conjunto de treino sendo este usado no treino de novo *ensemble* de cinco redes neuronais CPG. Este modelo permitiu uma correlação moderada ($R^2 = 0,600$) em ensaio de validação cruzada fraccionando o conjunto de treino em cinco subconjuntos. Estes resultados ilustram a dificuldade em encontrar descritores moleculares individualizados ou até mesmo um conjunto reduzido de descritores que permita a obtenção de um modelo robusto de previsão de pontos de fusão. Seguidamente esta discussão vai concentrar-se em exclusivo nos dois modelos treinados com o conjunto de 92 descritores (uma rede neuronal CPG e um *ensemble* de 5 redes neuronais CPG – **Tabelas II.26 e II.27**), já que foram os modelos que apresentaram melhor capacidade de previsão. A

arquitetura das redes neuronais CPG permite visualizar a relação entre estrutura do composto e a propriedade de interesse, neste caso o ponto de fusão. Nesse sentido vai ser usado nesta discussão um conjunto de 17 sais de guanidínio diversificados estruturalmente tanto ao nível do catião como do anião (**Figura II.27**), que foram mapeados em modelo baseado numa rede neuronal de *Counter-Propagation*. Na **Figura II.28** aparece o pormenor de cada uma das camadas de *output*, referentes a cada um dos quatro aniões considerados no modelo, onde são mapeados os 17 sais representados na **Figura II.27**, com o anião a ditar em qual das quatro camadas de *output* é mapeado cada sal. As cores dos neurónios das quatro camadas de *output* (**Figura II.28**) representam gamas de pesos dos neurónios de *output*. Os pesos de *output* representam valores de ponto de fusão produzidos como previsões do modelo CPG-NN quando os respectivos pesos são activados. Neurónios com cor azul representam elevados pontos de fusão, enquanto a coloração vermelha está associada a baixos pontos de fusão. Analisando as estruturas mapeadas em cada uma das camadas de *output* (**Figura II.28**) foi possível reconhecer certas regiões de pontos de fusão correspondendo a determinados tipos de estruturas. As cores associadas a diferentes regiões de pontos de fusão estão altamente correlacionadas nos quatro mapas estudados (**Figura II.28** – cada mapa corresponde a um tipo específico de anião), sugerindo que a relação entre ponto de fusão e estrutura de catião é a mesma para os quatro aniões estudados. Há que ter ainda em consideração o facto de cada camada de *output* ter topologia toroidal de modo que todos os neurónios têm exactamente o mesmo número de vizinhos – os neurónios da extremidade esquerda são adjacentes aos neurónios da extremidade direita, o mesmo em relação às extremidades inferior e superior (**Figura II.28**).

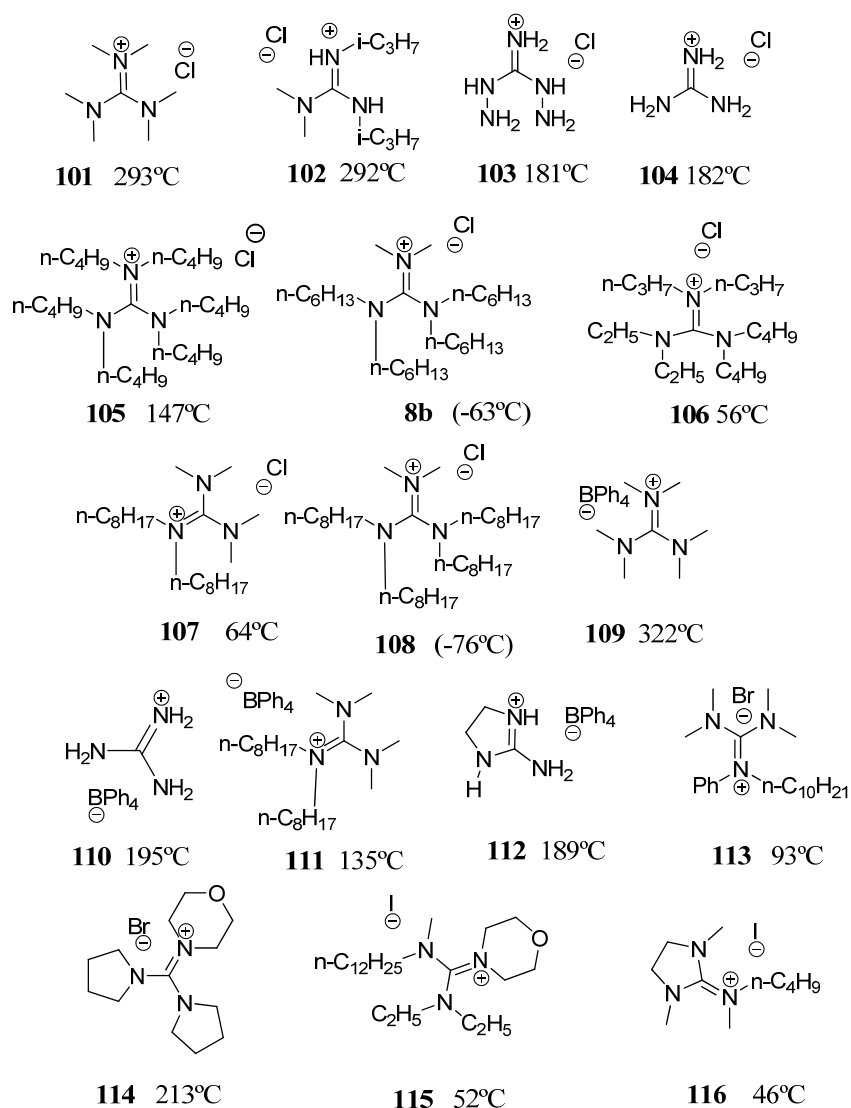


Figura II.27: Estruturas de sais de guanidínio mapeadas em modelo baseado numa rede de *Counter-Propagation*.

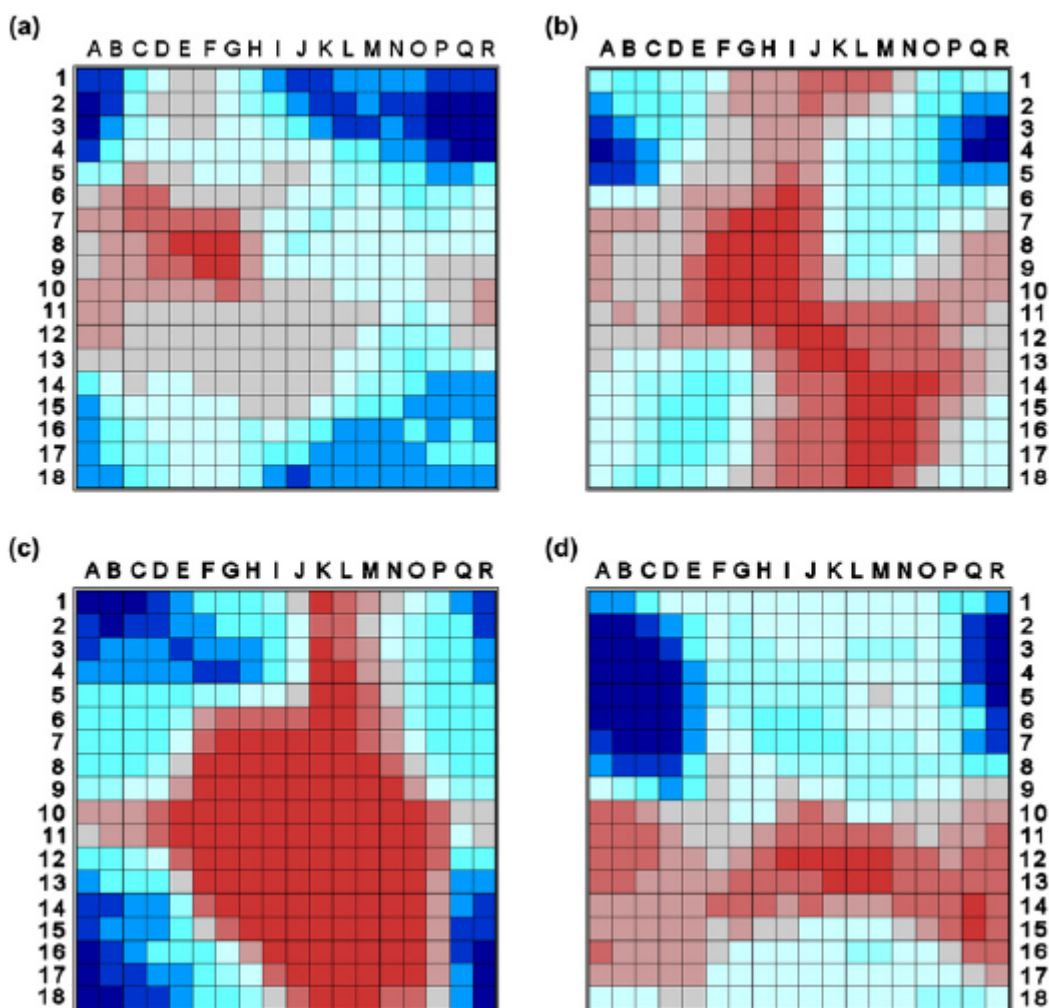


Figura II.28: Representação e comparação das quatro camadas de *output* de modelo baseado numa rede neuronal de *Counter-Propagation*, correspondendo a quatro famílias aniónicas, (a) cloreto (Cl^-), (b) tetrafenilborato (BPh_4^-), (c) brometo (Br^-) e (d) iodeto (I^-). As cores codificam os pesos da camada de *output* – neurónios azuis representam pontos de fusão elevados, enquanto neurónios vermelhos representam pontos de fusão baixos.

Foi observado que catiões a activar regiões correspondentes a neurónios de coloração azul escura (ponto de fusão muito elevado) correspondem a catiões de guanidínio contendo posições não substituídas e pequenos grupos alquílicos (**Figura II.14 – Estrutura 93**). A representar estas regiões estão os sais **101** e **109** (neurónio R3, **Figura II.28 “a”** e **“b”** respectivamente), o sal **102** (neurónio Q5, **Figura II.28 “a”**), o sal **103** (neurónio Q17, **Figura II.28 “a”**), os sais **104** e **110** (neurónio R18, **Figura II.28 “a”** e **“b”** respectivamente), o sal **112** (neurónio A16, **Figura II.28 “b”**) e o sal **114** (neurónio C18, **Figura II.28 “c”**). As estruturas moleculares dos exemplos mencionados estão representadas na **Figura II.27**. Regiões azul pálido, correspondendo a pontos de fusão intermédios estão tipicamente associadas a estruturas de guanidínios com distribuição simétrica de pequenas cadeias alquílicas. Como exemplo surge o cloreto de hexabutilguanidínio **105** (neurónio F5, **Figura**

II.28 “a”). As regiões correspondentes a cor vermelha (baixo ponto de fusão) foram activadas por estruturas de guanidínios incorporando em simultâneo cadeias longas e cadeias curtas, tal como a estrutura **8b** (neurónio F8, **Figura II.28 “a”**), estrutura **106** (neurónio B10, **Figura II.28 “a”**), estruturas **107** e **111** (neurónio G11, **Figura II.28 “a”** e **“b”** respectivamente), estrutura **113** (neurónio I14, **Figura II.28 “c”**), estrutura **115** (neurónio L12, **Figura II.28 “d”**) e a estrutura **116** (neurónio P13, **Figura II.28 “d”**).

A influência de determinadas características estruturais do catião guanidínio no ponto de fusão é, em geral, semelhante à influência exercida noutros tipos de catiões.⁸² Longas cadeias alquílicas provocam um decréscimo na importância de interações Coulombicas promovendo o aumento do número de graus conformacionais associados à estrutura (aumentando a entropia de fusão) provocando assim o decréscimo do ponto de fusão. No sentido inverso o aumento da simetria permite uma melhor acomodação na estrutura cristalina aumentando o ponto de fusão. Observou-se igualmente que os sais contendo anião cloreto têm tendência a apresentar pontos de fusão mais baixos que os seus homólogos contendo outro anião. Ainda assim, a distribuição das cores pelos diferentes mapas é muito similar, isto significa que, estruturas catiónicas correspondendo aos mais baixos (ou mais elevados) pontos de fusão para um determinado anião, vão igualmente corresponder aos mais baixos (ou mais elevados) pontos de fusão para outro anião. As principais diferenças nos quatro mapas (**Figura II.28**) estão relacionadas com as diferentes composições dos conjuntos de dados para os diferentes aniões, i.e., um catião pode estar disponível para determinado anião, mas não para outro anião – por exemplo a região superior direita da **Figura II.28 “d”**, está particularmente representada por sais de iodeto. Esta análise da relação entre estrutura e ponto de fusão baseou-se no modelo de uma rede neuronal de *Counter-Propagation* referido nas **Tabelas II.26** e **II.27**.

Depois de construir modelos para a previsão de pontos de fusão de sais de guanidínios e de seleccionar e validar o melhor modelo (*ensemble* de cinco redes neuronais CPG), este foi usado no *design* e síntese de novos sais de guanidínio. Concretamente, foram sintetizados seis sais (**Figura II.29**) - três cloretos (**4a**, **5a**, e **6a**) e três sais de tetrafenilborato (**7a**, **8a**, e **9a**) - após indicação do ponto de fusão por parte do modelo baseado em *ensemble* de cinco redes neuronais CPG.

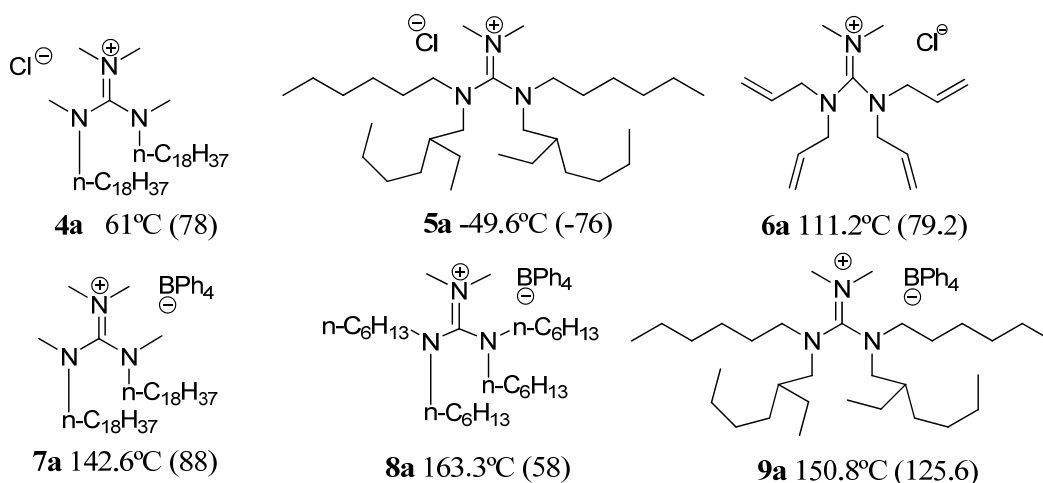
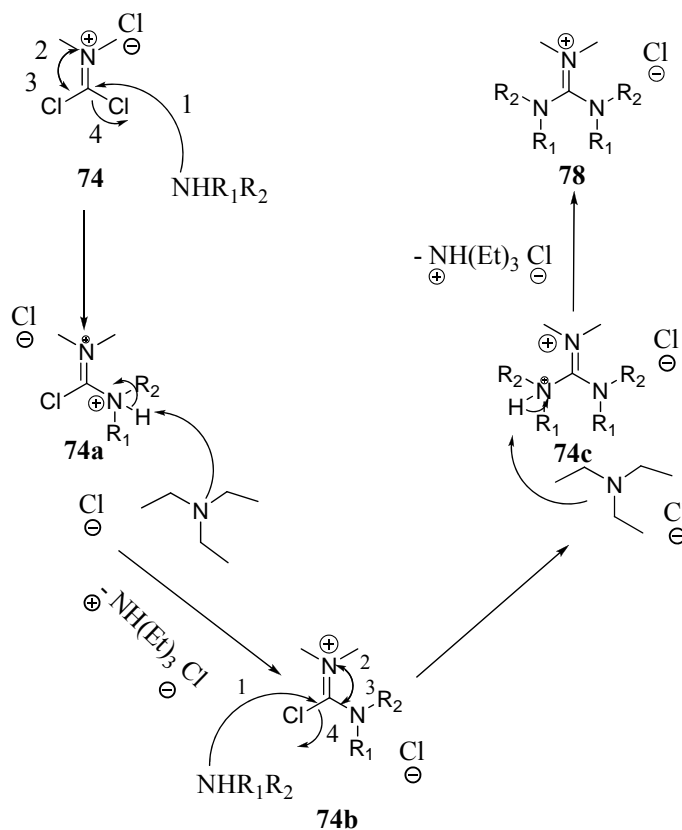


Figura II.29: Novos sais de guanidínio sintetizados, pontos de fusão previstos por modelo baseado em *ensemble* de cinco redes neuronais CPG e pontos de fusão experimentais (em parênteses).

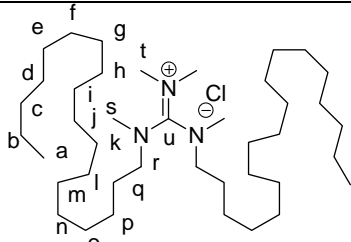
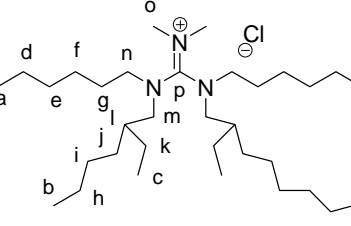
Os seis compostos em causa (**Figura II.29**) foram sintetizados de acordo com procedimento experimental⁴². Para a síntese destes sais seguiu-se primeiramente o método de síntese descrito no **Esquema II.14** a partir de cloreto de *N,N*-dimetilfosgenimínio (**74**) e amina secundária⁴² na presença de trietilamina. O mecanismo reaccional é o seguinte (**Esquema II.19**):



Esquema II.19: Mecanismo reaccional para a síntese de cloretos de guanidínio.

Segundo este mecanismo reaccional ocorre ataque nucleófilo por parte de uma amina secundária ao cloreto de *N,N*-dimetilfosgenimínio (**74**), formando-se **74a**. O derivado **74a** é desprotonado na presença de trietilamina (base) formando-se o intermediário **74b**, este é novamente atacado nucleofilicamente por amina secundária, resultando **74c**. Seguidamente ocorre de novo desprotonação na presença de trietilamina, resultando no final o cloreto de guanidínio **78**. Este método de síntese foi usado na síntese dos cloretos de guanidínio **4a**, **5a**, **6a** e **8b** representados na **Figura II.29**, **Tabelas II.29** e **II.30**.

Tabela II.29: Dados reaccionais, espectroscópicos e análises elementares dos cloretos de guanidínio sintetizados.

	Tempo de reacção η (%)	$^1\text{H-RMN}$ δ (ppm) J (Hz)	$^{13}\text{C-RMN}$ δ (ppm)	IV(filme ou KBr) $\nu_{\text{máx}}$ (cm^{-1})	Análise elementar
 4a 61°C (78)	20 horas 40%	0,85 (6H, t, $J = 6,8$) Ha 1,22 (64H, complexo) Hb-q 2,92-3,10 (16H, m) Hr, Hs, Ht	14,03 Ca 22,61 Cb 26,76 Cp 27,52 Cq 28,92 29,36, 29,46, 29,69 31,85 Cc-o 38,62, 38,86, 40,38 e 41,36 Cs,t 52,80 Cr 163,73 Cu	IV (KBr) 2920 (CH alifática) 2850 (CH alifática) 1593 (C-N) 1566(C-N)	composição calculada para $\text{C}_{41}\text{H}_{86}\text{ClN}_3 \cdot 1,5\text{H}_2\text{O}$ (%): N 6,15, C 72,03, H 13,12. composição determinada (%): N 6,24, C 72,06, H 13,45.
 5a -49,6°C (-76)	6 horas 64%	0,80-0,90 (18H, complexo) Ha,b,c 1,16-1,27 (28H, complexo) Hd-j 1,41-1,53 (4H, m) Hk 1,93-1,96 (2H, m) Hl 2,76 (6H, s) Ho 3,05 (4H, d, $J=7,2$) Hm 3,17 (4H, m) Hn	10,04, 10,45 Cc 13,83, 13,91, 13,96 Ca,b 22,41, 22,78, 23,58 e 23,71 Cd,h 26,41, 26,55 Ck 27,61, 28,14, 28,53 Cf 30,31, 30,46, 31,14, 31,45 Ce,g,i,j 35,88 Cl 38,70 Co 48,21, 48,68, 50,69, 51,05 Cm,n 165,17 Cp	IV (filme) 2927 (CH alifática) 2857 (CH alifática) 1643 (C-N) 1577 (C-N) 1535 (C-N)	composição calculada para $\text{C}_{31}\text{H}_{66}\text{ClN}_3 \cdot 1,25\text{H}_2\text{O}$ (%): N 7,80 C 69,10 H 12,81. composição determinada (%): N 7,52 C 68,96 H 13,35.

A acompanhar as estruturas aparece um código alfanumérico juntamente com o ponto de fusão previsto pelo modelo de previsão e o ponto de fusão ou transição vítrea experimentais em parênteses. η – rendimento (%), δ - desvio químico, J – constante de acoplamento, $\nu_{\text{máx}}$ – energia representada em número de onda.

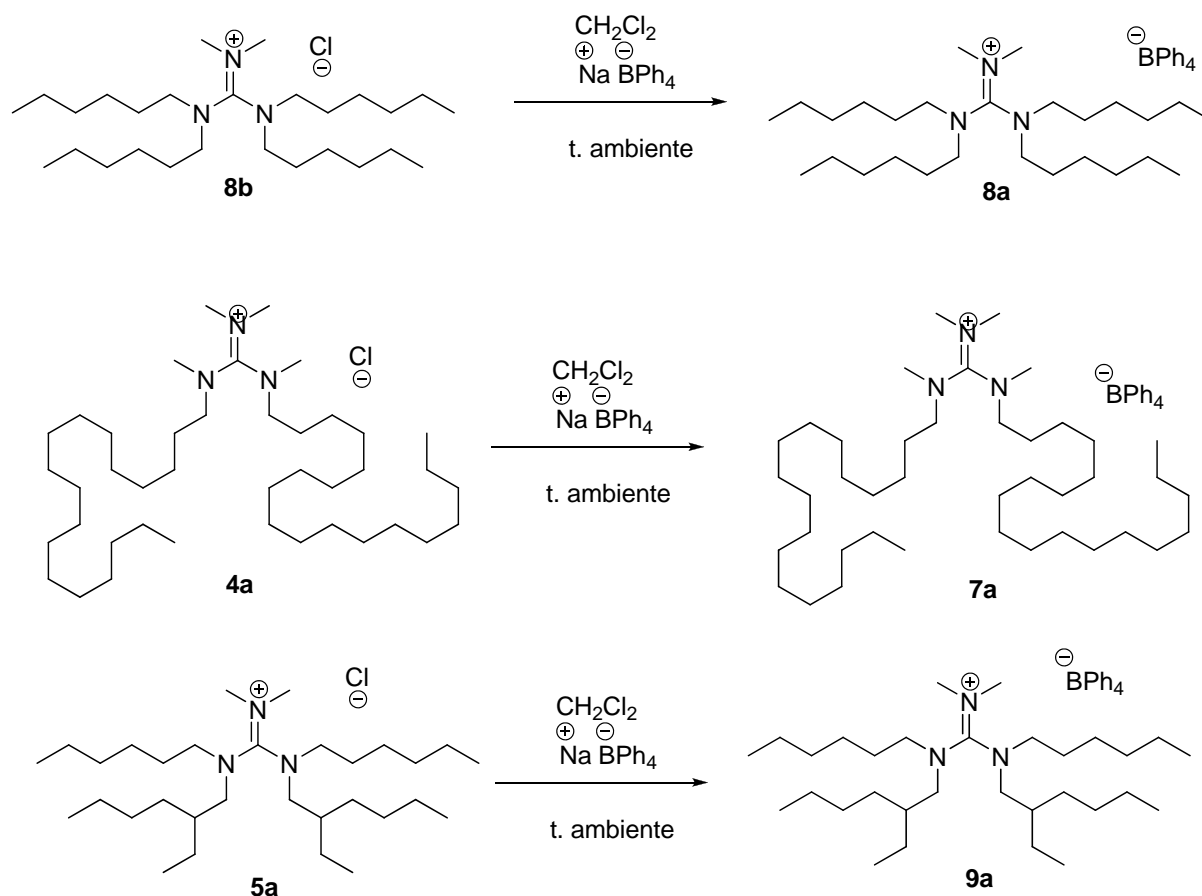
Tabela II.30: Dados reaccionais, espectroscópicos e análises elementares dos cloretos de guanidínio sintetizados.

	Tempo de reacção η (%)	^1H -RMN δ (ppm) J (Hz)	^{13}C -RMN δ (ppm)	IV(filme ou KBr) $\nu_{\text{máx}}$ (cm^{-1})	Análise elementar
 6a 111,2°C (79,2)	10 horas 65%	3,06 (6H, s), Ha 3,78-3,96 (8H, m) Hd 5,28-5,32 (8H, m) Hb 5,76 (4H, m) Hc.	40,97 Ca 52,64 e 52,89 Cd 121,57 Cb 131,11 Ce 163,03 Ce	IV (KBr) 3082 (CH olefínica) 2978 (CH alifática) 2920 (CH alifática) 1643 (C-N) 1589 (C-N) 1539 (C-N) 999 (CH olefínica) 933 (CH olefínica) 876 (CH olefínica)	composição calculada para $\text{C}_{15}\text{H}_{26}\text{ClN}_3 \cdot 1,7 \text{H}_2\text{O}$ (%): N 13,36, C 57,29, H 9,42. composição determinada (%): N 13,48, C 56,91, H 9,63.
 8b -29°C "objecto pertencente ao conjunto de treino" (-63°C) ⁴²	6 horas 43%	0,69-0,73 (12H, m) Ha 1,14 (32H, complexo) Hb-e 2,98 (6H, s) Hg 3,08-3,12 (8H, t, $J = 7,12$) Hf	13,53 e 13,58 Ca 22,11 Cb 26,10, 26,18, 26,28 Cd 27,20, 27,54 Ce 31,08 Cc 40,61 Cg 49,45 Cf 164,86 Ch	-	-

η – rendimento (%), δ - desvio químico, J – constante de acoplamento, $\nu_{\text{máx}}$ – energia representada em número de onda.

De acordo com as Tabelas II.29 e II.30 foram sintetizados quatro cloretos de guanidínio. O composto **6a**, foi sintetizado, caracterizado e usado na validação do modelo de previsão de pontos de fusão de sais de guanidínio baseado em *ensemble* de cinco redes neuronais CPG. Os compostos **4a** e **5a** foram usados para os mesmos fins que o composto **6a**, sendo aqueles adicionalmente usados como intermediários na síntese dos sais de tetrafenilborato **7a** e **9a** respectivamente. O cloreto **8b** foi usado exclusivamente como intermediário na síntese do sal de tetrafenilborato **8a**. Tendo em conta os dados espectrais e as análises elementares realizadas aos compostos presentes na Tabela II.29 e II.30 pode-se afirmar que os compostos pretendidos foram sintetizados com elevado grau de pureza, concretamente os dados dos espectros ^1H -RMN evidenciam uma grande correspondência entre os desvios químicos detectados e as funcionalidades presentes tal como a equiparação entre as áreas relativas a cada desvio químico e o número de prótons de determinado tipo presentes em cada composto. Característicos destes tipo de compostos são os desvios situados na gama 2,75-3,20 ppm típicos de prótons situados em carbonos adjacentes a azotos presentes no grupo guanidínio, a única excepção são os prótons Hd do composto **6a** que correspondem a desvio químico a 3,78-3,96 ppm pois também estão adjacentes a funcionalidade olefínica. A presença destes desvios químicos e as

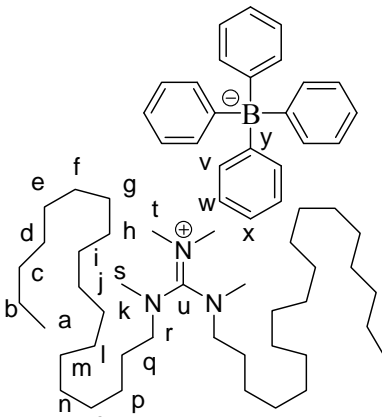
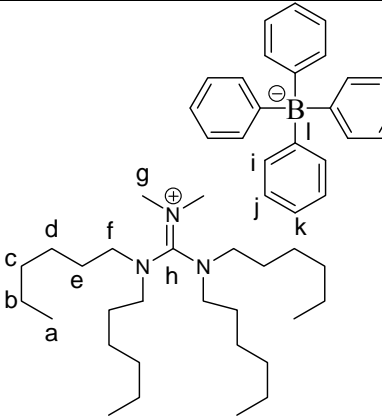
suas respectivas áreas evidenciam que a troca dos dois cloros, no cloreto de *N,N*-dialquilfosgenimínio por duas aminas secundárias, foi completa formando-se sal de guanidínio. Relativamente aos dados espectrais de ^{13}C -RMN, estes indiciam a presença do composto pretendido, em cada um dos quatro casos, e com elevado grau de pureza, neste caso há a destacar os desvios químicos a 163-165,5 ppm típicos de carbono quaternário presente no núcleo guanidínio. Os dados relativos às análises elementares confirmam a presença dos compostos pretendidos com elevado grau de pureza tal como indica a correspondência entre a composição calculada e a composição determinadas experimentalmente, estes dados revelam também a presença de água em quantidades significativas na proporção de 1,25-1,70 moléculas de água para 1 molécula de composto pretendido. De referir que após isolamento dos compostos da **Tabela II.29** e **II.30** estes foram submetidos a alto vácuo (10^{-6} mbar) durante *ca* de 4h, e é curioso que apesar deste procedimento os cloretos sintetizados indicavam a presença de água em quantidades muito significativas. Apesar das quantidades de água evidenciadas, os compostos **4a** e **6a** foram isolados como sólidos, o mesmo não acontecendo com os compostos **5a** e **8b** que foram obtidos como líquidos. Ainda assim há que olhar para os valores de ponto de fusão e transição vítrea com a devida precaução dada a presença de água. Completando a análise aos dados de infravermelho, estes apresentam bandas características de ligações C-N de guanidínios ($1650\text{-}1500\text{ cm}^{-1}$) tal como bandas a $2850\text{-}3000\text{ cm}^{-1}$ típicas de ligações C-H de grupos alquílicos no caso concreto dos compostos **4a** e **5a**, no caso exclusivo do composto **6a** aparece banda a 3082 cm^{-1} e bandas entre $850\text{-}1000\text{ cm}^{-1}$ típicas de ligações C-H olefínicas. O composto **8b** não foi caracterizado por espectroscopia de Infravermelho nem por análise elementar dado ser em exclusivo um composto intermediário na síntese do sal de tetrafenilborato **8a**. Da mesma forma os compostos **4a** e **5a** foram também intermediários na síntese dos sais de tetrafenilborato **7a** e **9a** respectivamente e de acordo com o seguinte esquema de síntese (**Esquema II.20**):



Esquema II.20: Reações de troca de anião cloreto por tetrafenilborato.

Os três sais de tetrafenilborato sintetizados foram caracterizados espectroscopicamente e por análise elementar (**Tabela II.31**) de forma a avaliar se a troca de anião foi completa (catião e anião pretendido na proporção 1:1) e se cada composto sintetizado está puro.

Tabela II.31-1: Dados reaccionais, espectroscópicos e análises elementares dos sais de guanidínio sintetizados cujo anião é o tetrafenilborato.

	Tempo de reacção η (%)	^1H -RMN δ (ppm) J (Hz)	^{13}C -RMN δ (ppm)	IV(filme ou KBr) $\nu_{\text{máx}}$ (cm^{-1})	Análise elementar
 <p>7a 142,6°C (88)</p>	24 horas 59,7%	0,89 (6H, t, $J = 6,92$) Ha 1,27 (64H, complexo) Hb-q 2,48 (12H, complexo) Hs,t 2,81 (4H, m) Hr 6,89 (4H, t, $J = 6,92$) Hx 7,03 (8H, t, $J = 7,2$) Hv 7,42 (8H, m) Hw	14,07 Ca 22,65 Cb 26,68 Cp, 27,47 Cq 29,17, 29,32, 29,55, 29,67, 31,88 37,68, 39,91, 40,20 Cs,t 52,62 Cr 121,69, 125,51, 136,17 Cv-x 163,00 Cu 163,34 163,83, 164,32, 164,81 Cy	IV (KBr) 3051 (CH aromática) 2920 (CH alifática) 2850 (CH alifática), 1566 (C-N), 1466 e 1408 (C=C aromática)	composição calculada para $\text{C}_{65}\text{H}_{106}\text{BN}_3$ (%): N 4,47, C 83,02, H 11,36. composição determinada (%): N 4,53, C 82,73, H 11,63.
 <p>8a 163,3°C (58)</p>	24 horas 62%	0,83-0,92 (12H, t, $J = 6,8$) Ha 1,19-1,32 (32H, complexo) Hb-e 2,21 (6H, s) Hg 2,63-2,97 (8H, m) Hf 6,91 (4H, t, $J = 7,16$) Hk 7,06 (8H, t, $J = 7,2$) Hi 7,48 (8H, m) Hj	13,88 Ca 22,32 Cb 26,62 Cd 27,32, 27,83 Ce 31,25, 31,33 Cc 39,78 Cg 49,29, 50,10 Cf 121,71, 125,64, 136,12 Ci-k 162,79 Ch 163,40, 163,89, 164,38, 164,87 Cl	IV (KBr) 3055 (CH aromática) 2927 e 2858 (CH alifática) 1574 (C-N) 1543 (C-N) 1462 (C=C aromática)	composição calculada para $\text{C}_{51}\text{H}_{78}\text{BN}_3 \cdot 0,1 \text{H}_2\text{O}$ (%): N 5,63, C 82,13, H 10,57. composição determinada (%): N 5,59, C 82,00, H 10,68.

η – rendimento (%), δ - desvio químico, J – constante de acoplamento, $\nu_{\text{máx}}$ – energia representada em número de onda.

Tabela II.31-2: Continuação da apresentação dos dados reaccionais, espectroscópicos e análises elementares dos sais de guanidínio sintetizados cujo anião é o tetrafenilborato.

[illegible]

η – rendimento (%), δ - desvio químico, J – constante de acoplamento, $\nu_{m\acute{a}x}$ – energia representada em número de onda.

Analisando os espectros de ^1H -RMN pode-se afirmar que as três reacções de troca de anião (**Esquema II.20**) foram completas já que as áreas relativas aos prótons aromáticos do anião (na gama 6,89-7,79 ppm) correspondem às áreas dos prótons do catião, tendo em conta o número de prótons de cada tipo e a proporção entre catião e anião ser 1:1. No caso do composto **7a** pode-se comparar a área de **Ha** e a área de **Hx**, relativamente ao composto **8a** pode-se comparar a área de **Ha** com **Hk**, finalmente no caso de **9a** pode-se comparar **Ho** com **Hs** (**Tabela II.31**). A presença dos catiões e aniões pretendidos é confirmada pelos espectros de ^1H -RMN. Relativamente aos espectros de ^{13}C -

RMN confirmam a presença dos catiões e aniões pretendidos, para ilustrar esse facto, e no caso do catião há que destacar o carbono quaternário do núcleo guanidínio perto de 163 ppm (Cu e Ch em **7a** e **8a** respectivamente), no caso de **9a** não foi possível identificar o sinal do carbono quaternário sendo este um sinal na generalidade muito pouco intenso pelo facto de corresponder a um único carbono (por unidade de composto) e o carbono ser quaternário, apesar desse facto foi possível identificar todos os outros carbonos do catião correspondendo aos desvios químicos esperados. Relativamente ao anião foi possível identificar todos os carbonos aromáticos tanto o carbono adjacente ao Boro (a *ca* de 165 ppm) tal como os outros carbonos aromáticos (115-136 ppm).

Relativamente aos espectros de Infravermelho em todos os casos foi possível identificar todos os tipos de funcionalidades presentes nos compostos alvo, ligações C-H aromáticas do anião perto de 3050 cm⁻¹, ligações C-H alifáticas do catião na gama 2850-2930 cm⁻¹, ligações C-N do catião guanidínio na gama 1500-1580 cm⁻¹ e ligações C-C aromáticas do anião na gama 1400-1470 cm⁻¹. Relativamente às análises elementares, e para o composto **9a**, verifica-se a presença de água sendo necessário adicionar 0,9 equivalentes de água à fórmula do composto **9a** para que a composição calculada coincidissem com a composição determinada experimentalmente. Relativamente a **7a** e **8a** a proporção de água necessária, para explicar as respectivas composições determinadas experimentalmente, foi nula ou negligenciável. Em todos os casos os resultados das análises elementares confirmam a pureza dos três sais de tetrafenilborato sintetizados.

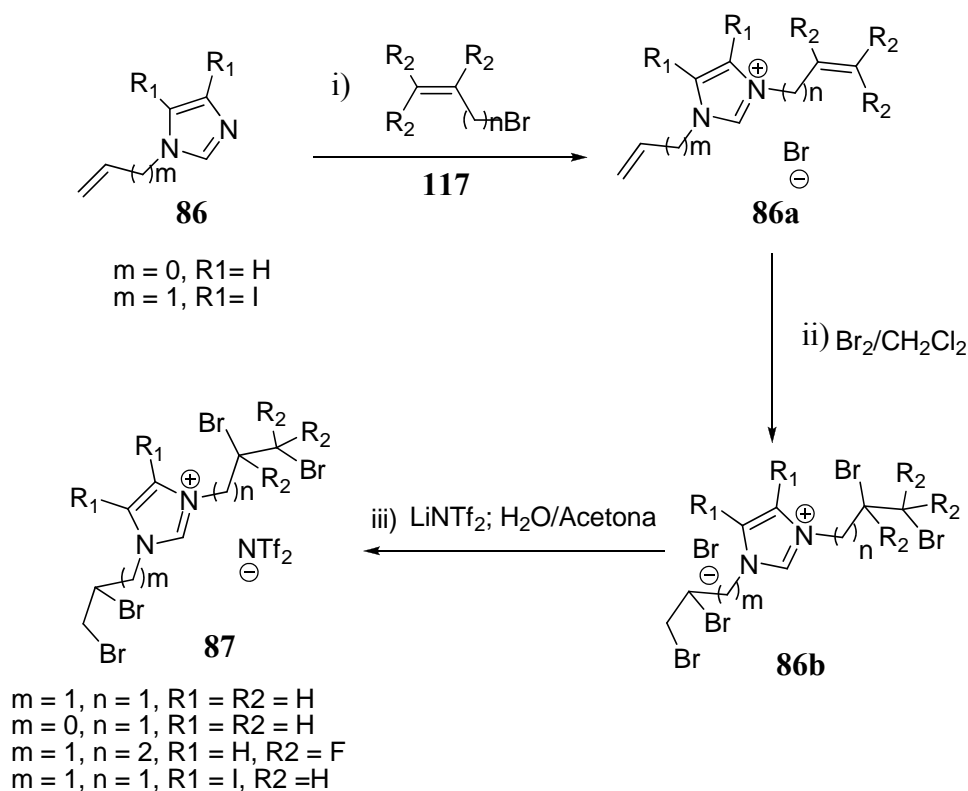
Os três sais de tetrafenilborato sintetizados são sólidos à temperatura ambiente, ainda assim há que olhar com a devida precaução para o ponto de fusão experimental do composto **9a** dada a presença de água em quantidade significativa.

Os catiões correspondentes às estruturas **5a**, **8a** e **9a** activam o neurónio vermelho F8 (**Figura II.28** – “a” no caso do cloreto **5a** e “b” no caso dos sais de tetrafenilborato **8a** e **9a**), este neurónio que corresponde a baixo ponto de fusão é também activado por estruturas **108** (Tg = -76°C) e **8b** (Tg = -63°C). Estas três estruturas catiónicas possuem quatro longos grupos alquílicos e dois grupos metilo substituintes. Os compostos **4a**, **7a** (activam neurónio J8, **Figura II.28** “a” e “b” respectivamente) e **6a** (activa neurónio Q7, **Figura II.28** “a”) activam neurónios em regiões de pontos de fusão intermédios. Enquanto que as estruturas **4a** e **7a** possuem quatro grupos metilo e dois grupos octadecilo substituintes, por sua vez a estrutura **6a** é muito simétrica com dois grupos metilo e quatro cadeias C3 com ligação dupla substituintes. Depois destes seis compostos (**4a**, **5a**, **6a**, **7a**, **8a** e **9a**) terem sido sintetizados foi medido o respectivo ponto de fusão ou transição vítrea, estes valores foram comparados com as previsões fornecidas pelo modelo de previsão de pontos de fusão baseado em *ensemble* de 5 redes neuronais CPG. Relativamente aos três cloretos sintetizados (**4a**, **5a** e **6a**) há uma boa correspondência entre pontos de fusão (ou transição vítrea) experimentais e pontos de fusão previstos (**Figura II.29**) correspondendo a um erro médio de 25°C, com **5a** a apresentar um valor

experimental de transição vítrea baixo (-76°C – **Figura II.29**), por sua vez **4a** e **6a** apresentam pontos de fusão intermédios. Nestes três casos há que olhar com a devida precaução para os pontos de fusão (ou transição vítrea) experimentais dada a quantidade significativa de água presente em qualquer um deles. Adicionalmente a estrutura **4a** activa neurónio numa região muito instável com poucos neurónios ocupados, logo a previsão obtida não é muito credível. Por outro lado o ponto de fusão mais baixo no conjunto de dados relativamente a sais de tetrafenilborato é 127°C , neste contexto é impossível obter previsões inferiores a este valor. Tendo em consideração este facto e observando que os pontos de fusão experimentais dos três sais de tetrafenilborato sintetizados (**7a**, **8a** e **9a**) são inferiores a este limite, pode concluir-se que este modelo ou outro construído com o mesmo conjunto de objectos nunca poderia acertar nos pontos de fusão experimentais destes três compostos ainda assim as previsões deste modelo estão muito próximas do limite mínimo de 127°C já que todas estão contidas numa gama de 36°C abaixo deste limite (**Figura II.29**). Também aqui há que ter em conta o facto de o composto **9a** possuir uma quantidade significativa de água.

II-3.5 - Estudo E: Síntese e Estudo de Sais de Guanidínio de Elevada Densidade

Um dos objectivos do estudo anterior era a obtenção de novos sais de guanidínio com baixo ponto de fusão, esse objectivo foi concretizado através do recurso a modelos de previsão de ponto de fusão baseado em redes de *Counter-Propagation*. No estudo desta secção pretende-se obter novos sais de guanidínio de elevada densidade mas dada a escassa informação presente na literatura relativamente à densidade de sais de guanidínio não pôde ser seguida a mesma abordagem da secção anterior de construir um modelo que indicasse novos sais de guanidínio densos. Este é um estudo exclusivo em síntese orgânica baseado no estudo de Ye e Shreeve,⁹² nesse estudo foram sintetizados novos líquidos iónicos de elevada densidade baseados na unidade imidazólio contendo número variado de halogénios pesados como é o caso do bromo e o iodo, alguns destes compostos foram preparados segundo o seguinte esquema de síntese (**Esquema II.21**):



Esquema II.21: Método sintético usado na síntese de sais de imidazólio bromados de elevada densidade. Bromação de posições olefinicas.

Neste Esquema de síntese, na primeira fase, alquilou-se o 1-alilimidazólio **86** com 3-bromopropeno sendo obtido o sal brometo **86a**. Na segunda fase o sal obtido (**86a**) foi bromado na presença de bromo (Br_2), na terceira e última fase o sal brometo resultante **86b** foi sujeito a reacção de troca de anião na presença de bis(trifluorometanosulfonil)imidato de lítio de forma a obter líquidos iónicos à temperatura ambiente do mesmo tipo representado na estrutura **87**. Seguindo o mesmo princípio de bromar posições olefinicas para a obtenção de um composto denso, procurou-se neste estudo obter sais de guanidínio bromados densos. Nesta transposição há que ter em consideração que o catião guanidínio tem mais grupos substituintes passíveis de conter posições olefinicas passíveis de serem bromadas do que imidazólios 1,3-disubstituídos (**Figura II.30**):

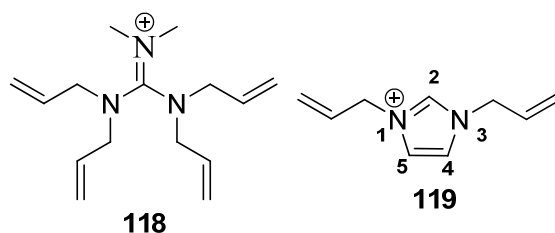
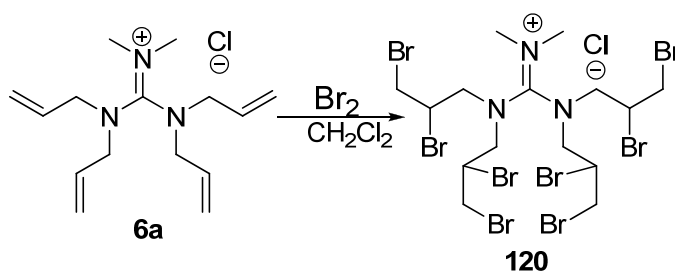


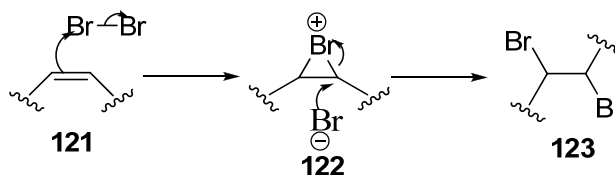
Figura II.30: Comparação relativamente ao número de grupos substituintes contendo funcionalidades olefinicas passíveis de serem bromadas entre catiões guanidínio e imidazólio 1,3 disubstituído.

Neste contexto parece possível a obtenção de novos compostos densos a partir de sais de guanidínio. Tendo em conta que no estudo II-3.4. d) já havia sido sintetizado o sal **6a**, cloreto de *N,N*-dimetil-*N',N'',N'''*-tetraalilguanidínio, foi feita uma tentativa de bromar este composto de acordo com Esquema II.22:



Esquema II.22: Método A - Reacção de bromação do cloreto de *N,N*-dimetil-*N',N'',N'''*-tetraalilguanidínio **6a** para a obtenção do cloreto de *N,N*-dimetil-*N',N'',N'''*-tetraquis(2,3-dibromopropil) guanidínio **120**.

Para esta primeira tentativa foi tido em consideração o mecanismo de bromação de posições olefinicas na presença de bromo (Br_2):



Esquema II.23: Mecanismo reaccional da bromação de posição olefinica.

Assim, como resultado do ataque nucleofílico de olefina **121** ao bromo forma-se o intermediário bromónio **122** que sofre ataque *anti* de anião brometo resultando no composto bromado **123**. Tendo em conta o mecanismo apresentado no Esquema II.23 e o composto de partida para esta síntese ser o cloreto **6a**, observou-se a possibilidade de ataque nucleofílico do anião cloreto, presente

no composto de partida **6a**, ao intermediário bromónio do tipo **122** em competição com o brometo, se esta possibilidade se concretizasse formar-se-ia uma mistura de compostos já que o composto de partida **6a** possui quatro posições olefinicas, neste contexto, os intermediários bromónios resultantes podem ser atacados tanto por cloreto como por brometo. De forma a demonstrar esta possibilidade foi realizada reacção de bromação tal como aparece no **Esquema II.22 – Método A**, findo período reaccional de 7h. a mistura resultante foi sujeita a reacção de troca de anião na presença de hexafluorofosfato de potássio durante 24h., no final deste período procedeu-se ao *work-up* da reacção em que a mistura reaccional foi filtrada e posteriormente a fase líquida foi lavada com sulfito de sódio e água de forma a eliminar o excesso de bromo, no final foi isolado um sólido amarelo que foi analisado por espectrometria de massa de forma a demonstrar a possibilidade de formação de múltiplos produtos resultado da competição entre cloreto e brometo para atacar os intermediários bromónio, o resultado evidenciado pelo espectro de massa foi esclarecedor (**Figura II.31** combinado com **Figura II.32**):

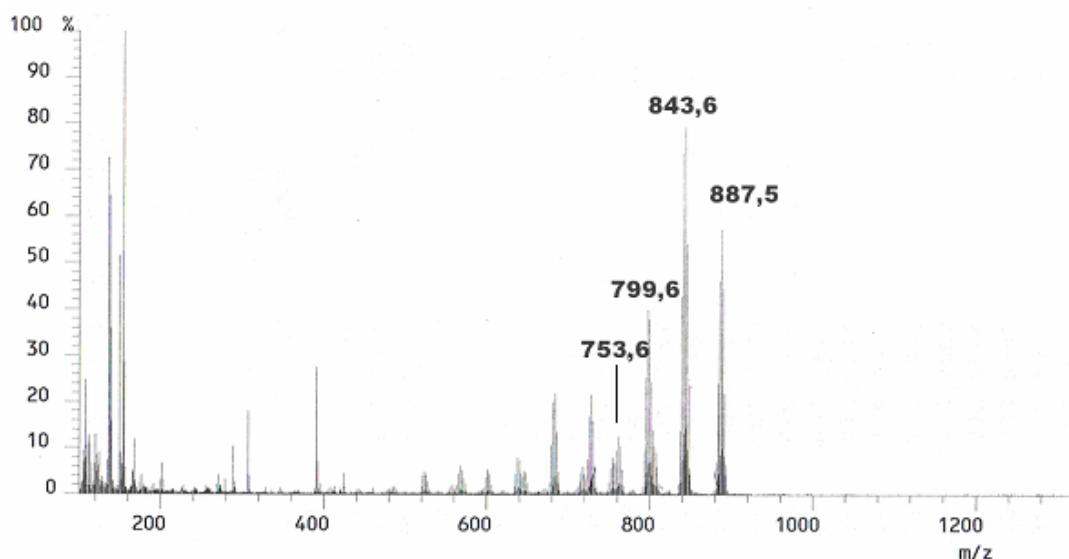


Figura II.31: Espectro de massa (FAB+) de sólido isolado de reacções de bromação e troca de anião partindo de cloreto de *N,N*-dimetil-*N',N',N'',N''*-tetraalilguanidínio **6a**. Aparece assinalada no topo do respectivo sinal a combinação maioritária de isótopos relativamente às estruturas apresentadas na **Figura II.32**.

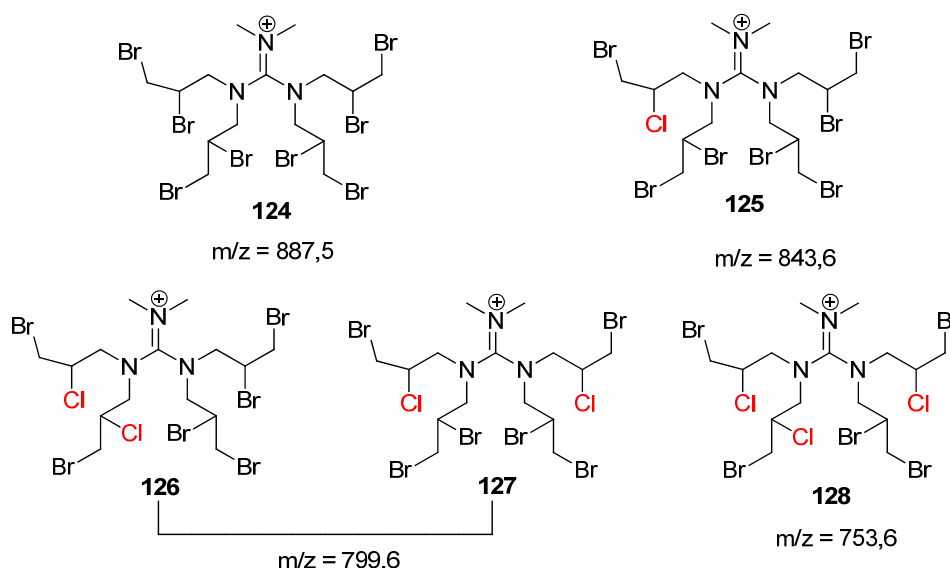
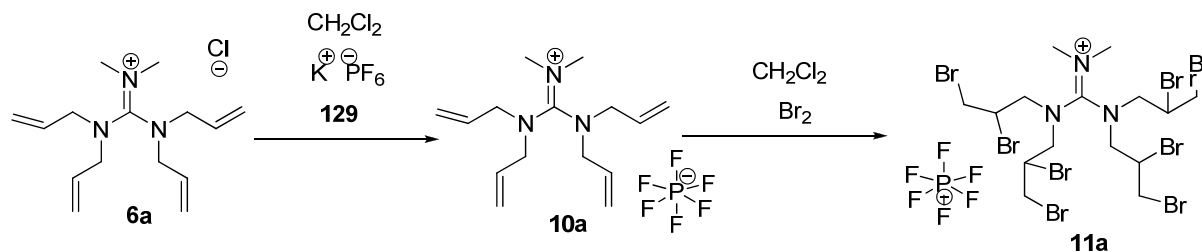


Figura II.32: Estruturas catiónicas correspondentes aos sinais indicados na Figura II.31.

Analisando o espectro de massa realizado ao sólido isolado (**Figura II.31**) é possível observar aglomerados de sinais em vez de sinais isolados, esta abundância de sinais em determinadas zonas do espectro deve-se ao facto de que o bromo existe na forma de dois isótopos estáveis (^{79}Br e ^{81}Br) com abundância praticamente equivalente, já o cloro também existe como dois isótopos estáveis (^{35}Cl e ^{37}Cl) em abundância de 75,77% e 24,23% respectivamente. Desta forma um determinado fragmento detectado por espectrometria de massa pode estar formado por diversas combinações de isótopos resultando numa multiplicidade de sinais muito próximos uns dos outros, mesmo assim em cada caso foi possível identificar a combinação maioritária representada por número em cima do respectivo sinal na **Figura II.31**. Tendo em conta esse facto foi possível reconhecer diversos tipos de catiões distintos correspondendo cada um a determinado sinal associado a um valor da relação massa/carga ou m/z (na prática cada valor em cima de sinal corresponde ao peso molecular de determinado catião). Neste caso o valor 887,5 (**Figura II.31** e **II.32**) corresponde ao peso molecular do catião totalmente bromado (**124 - Figura II.32**), o valor 843,6 corresponde ao catião com um átomo de cloro (**125 - Figura II.32**), o valor 799,6 corresponde aos catiões com dois átomos de cloro (**126 e 127 - Figura II.32**), e o valor 753,6 corresponde ao catião com três átomos de cloro (**128 - Figura II.32**). Fica assim demonstrado que não é boa ideia bromar sais de guanidínio quando o anião é nucleófilo como é o caso do cloreto, pelo menos no caso de se querer isolar um único produto. Desta forma foi necessário modificar o procedimento descrito por Ye e Shreeve⁹² para a síntese de sais de imidazólio densos, como no caso desse trabalho se partia de brometo de imidazólio podia-se realizar a bromação inicialmente sem o perigo de haver competição por parte de elementos diferentes para o ataque aos intermediários bromónio já que o anião brometo se baseava no mesmo elemento presente no bromo diatómico usado na bromação. No caso deste trabalho como o composto de partida **6a** é cloreto não é possível realizar em primeiro lugar a bromação e obter um produto isolado. De

forma a resolver esta situação, num segundo ensaio, decidiu-se submeter inicialmente **6a** à troca de anião de forma a obter um sal de guanidínio contendo anião não nucleófilo, o hexafluorofosfato, esse sal posteriormente foi submetido a reacção de bromação com o objectivo de obter sal de guanidínio de elevada densidade **11a** (**Esquema II.24**).



Esquema II.24: Método B - Método de síntese modificado para a obtenção de sal de guanidínio de elevada densidade 11a, baseado em bromação de posições olefinicas posteriormente à reacção de troca de cloreto por anião não nucleófilo.

No final da síntese de dois passos representada no **Esquema II.24 – Método B** foi obtido um sólido branco que foi analisado por espectrometria de massa sendo obtido o seguinte resultado (**Figura II.33**):

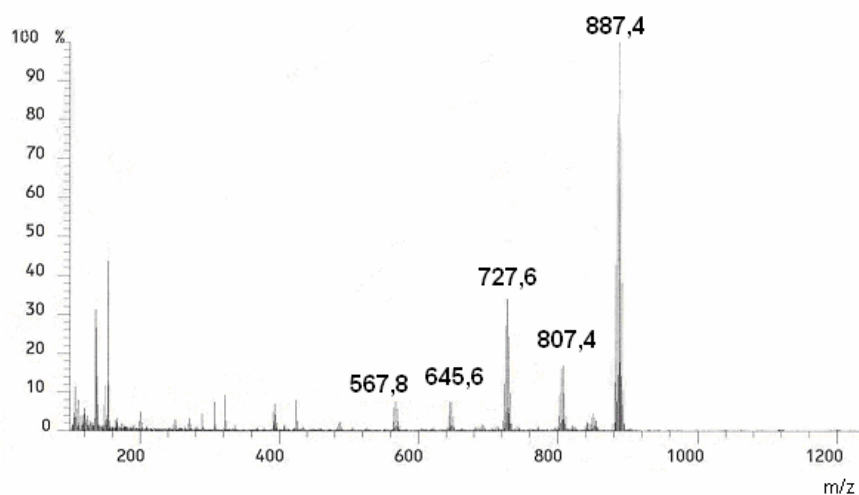
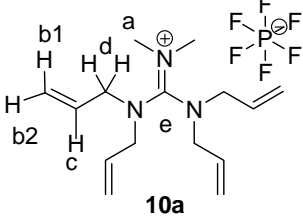


Figura II.33: Espectro de massa (FAB+) de sólido isolado resultado de método de síntese modificado realizando sequencialmente reacção de troca de anião e posteriormente reacção de bromação. Aparece assinalada no topo do respectivo sinal a combinação maioritária de isótopos relativamente aos fragmentos obtidos a partir da estrutura 11a.

O espectro de massa da **Figura II.33** apresenta sinal a $m/z = 887,4$ indicando a presença de catião totalmente bromado **124** (**Figura II.32**), adicionalmente pode-se observar que não aparecem sinais a $m/z = 843,6$, $799,6$ e $753,6$, tal como aparece no espectro de massa da **Figura II.31**, este facto indica a presença de apenas um produto contendo catião totalmente bromado **124**. Adicionalmente pode-se verificar sinais correspondendo a $m/z = 807,4$, $727,6$, $645,6$ e $567,8$ correspondendo a fragmentos resultantes da quebra de ligação carbono-bromo, indicando respectivamente a quebra de uma, duas, três e quatro ligações carbono-bromo no processo de obtenção do espectro de massa (**Figura II.33**). Ficou assim demonstrado que o método sintético exemplificado no **Esquema II.24** é adequado para a obtenção de um sal de guanidínio bromado, potencial material de elevada densidade. Adicionalmente o composto intermediário **10a** e o produto final **11a**, resultado da síntese representada no **Esquema II.24 – Método B**, foram caracterizados por ressonância magnética nuclear (^1H e ^{13}C), espectroscopia de infravermelho e análise elementar ou espectrometria de massa:

Tabela II.32: Dados reaccionais, espectrais e análise elementar do composto intermediário 10a à síntese de sal de guanidínio de elevada densidade 18a.

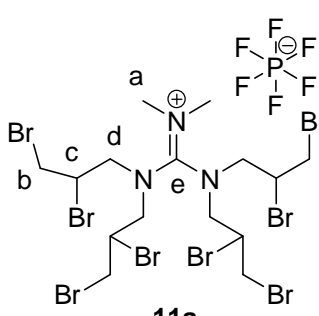
	Tempo de reacção η (%)	^1H -RMN δ (ppm) J (Hz)	^{13}C -RMN δ (ppm)	IV(filme ou KBr) $\nu_{\text{máx}}$ (cm^{-1})	Análise elementar
 <p>10a</p>	24 horas 89%	2,87 (6H, s) Ha 3,68 (8H, sl) Hd 5,25-5,29 (8H, m) Hb 5,66 (4H, m) Hc	39,94 Ca 52,03, 52,49 Cd 121,25, 121,46 Cb 130,63 Cc 162,69 Ce	IV (KBr) 3088 (CH olefínica) 3032 (CH olefínica) 2988 (CH alifática) 2945 (CH alifática) 2912 (CH alifática) 1645 (C-N) 1583 (C-N) 1535 (C-N) 1002 (CH olefínica) 956 (CH olefínica) 929 (CH olefínica) 837 (PF ₆)	Composição calculada para C₁₅H₂₆N₃F₆P (%) : N: 10,68 C: 45,80, H 6,66. composição determinada (%) : N 10,62, C 45,96, H 6,60.

η – rendimento (%), δ - desvio químico, J – constante de acoplamento, $\nu_{\text{máx}}$ – energia representada em número de onda.

Percorrendo a **Tabela II.32** é possível observar por análise a espectros de ^1H -RMN e ^{13}C -RMN que o catião pretendido *N,N*-dimetil-*N',N',N'',N''*-tetraalilguanidínio está presente no composto isolado já que todos os carbonos e hidrogénios do catião vão corresponder a sinais do espectro de ^1H -RMN e ^{13}C -RMN respectivamente. A destacar no espectro de ^1H -RMN o sinal correspondente aos protões Ha a 2,87 ppm e o sinal correspondente aos protões Hc a 5,66 ppm, ambos os sinais possuem área correspondente a 6 e 4 hidrogénios respectivamente indicando que a síntese do catião guanidínio foi

completa (**Esquema II.19**). Relativamente ao espectro de ^{13}C -RMN há a destacar o sinal a 162,69 ppm correspondendo ao carbono quaternário Ce característico de catiões guanidínio. Relativamente ao espectro de infravermelho, este confirma a presença do catião pretendido já que apresenta bandas características de ligações C-H de olefinas (3088 e 3032 cm^{-1}) tal como ligações C-H alifáticas (2988 e 2912 cm^{-1}), adicionalmente aparecem bandas características de ligações C-N de guanidínios (1645, 1583 e 1535 cm^{-1}). Toda esta informação confirma a presença do catião pretendido; relativamente ao anião, o espectro de infravermelho apresenta uma banda a 837 cm^{-1} característica de anião hexafluorofosfato. Se pudesse ainda haver algumas dúvidas relativamente à presença do catião e anião pretendidos no composto **10a** estas foram completamente dissipadas por análise elementar havendo correspondência entre composição determinada e composição calculada para a combinação de catião e anião pretendidos. O composto **10a** obtido não evidencia a presença de água. O composto **10a** foi posteriormente bromado, de acordo com **Esquema II.24**, sendo obtido um produto pretendendo-se que este corresponda à estrutura de **11a**. Este produto foi caracterizado por espectrometria de ^1H -RMN, ^{13}C -RMN, Infravermelho e espectrometria de massa (**Tabela II.33**).

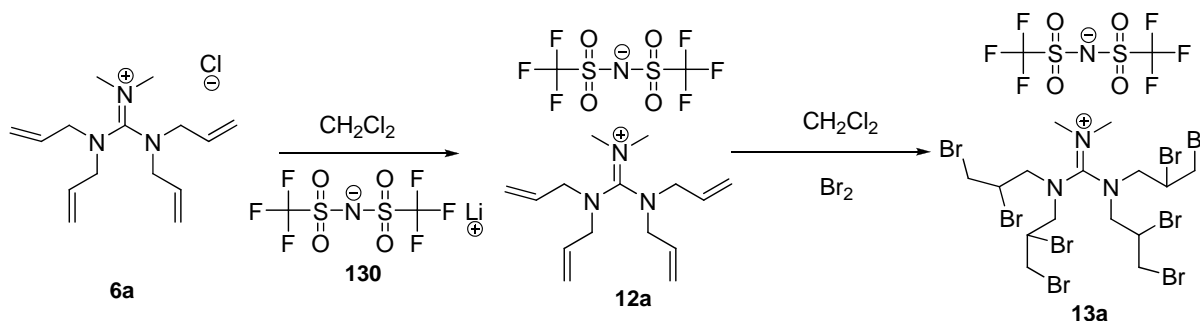
Tabela II.33: Dados reaccionais e espectrais de sal de guanidínio bromado 11a, nos sinais relativos ao espectro de massa (+) representa os fragmentos correspondentes à combinação de isótopos maioritária.

	Tempo de reacção η (%)	^1H -RMN δ (ppm) J (Hz)	^{13}C -RMN δ (ppm)	IV(filme ou KBr) ν_{\max} (cm^{-1})	Espectro de massa (FAB+)
	24 horas 13,8%	3,30-3,38 (8H, m) Hd 3,79-3,88 (14H, complexo) Ha,b 4,04-4,16 (4H, m) Hc	33,89, 33,94, 34,48, 34,69, 35,27, 35,38, 35,58, 35,77 Ca 42,18, 42,27, 42,32, 42,53, 42,63 Cb 46,43, 46,62, 47,32, 47,56, 47,62, 47,95, 48,09, 48,93 Cc 53,20, 53,35, 53,77, 54,31, 54,69, 54,78, 56,31, 57,91 Cd 163,89, 164,89 Ce	IV (KBr) 3001 (CH alifática) 2935 (CH alifática) 1585 (C-N) 1517 (C-N) 840 (PF ₆) 655 (C-Br) 575 (C-Br)	m/z 894 (M ⁺ , 2,1) 893 (M ⁺ , 10,7) 892 (M ⁺ , 6,7) 891 (M ⁺ , 38,3) 890 (M ⁺ , 14,2) 889 (M ⁺ , 78,5) 888 (M ⁺ , 17,8) 887 (M ⁺ , 100) + 885 (M ⁺ , 81,4) 884 (M ⁺ , 7,6) 883 (M ⁺ , 42,9) 882 (M ⁺ , 2,1) 881 (M ⁺ , 12,8) 807 (C ₁₅ H ₂₇ Br ₇ N ₃ ⁺ , 16,8) + 727 (C ₁₅ H ₂₆ Br ₆ N ₃ ⁺ , 34,1) + 645 (C ₁₅ H ₂₇ Br ₅ N ₃ ⁺ , 7,6) + 568 (C ₁₅ H ₂₆ Br ₄ N ₃ ⁺ , 7,7) +

η – rendimento (%), δ - desvio químico, J – constante de acoplamento, ν_{\max} – energia representada em número de onda, m/z relação massa/carga, FAB – Fast Atom Bombardment.

De acordo com os dados da Tabela II.33 é possível observar que no espectro de ^1H -RMN há correspondência entre os desvios químicos e respectivas áreas relativamente ao tipo e número de prótons presentes no catião bromado da estrutura 11a. Na conversão de 10a em 11a os prótons Ha ficam mais desblindados aparecendo em 11a no intervalo 3,79-3,88 ppm. Por sua vez Hb, Hc e Hd aparecem mais blindados em 11a correspondendo aos desvios nos intervalos 3,79-3,88, 4,04-4,16 e 3,30-3,38 ppm respectivamente. Relativamente ao espectro de ^{13}C -RMN há correspondência entre os sinais apresentados e os respectivos carbonos do catião da estrutura 11a pretendida. Na conversão de 10a em 11a o carbono Ca fica mais blindado (33,89-35,57 ppm), Cb fica muitíssimo mais blindado (42,18-42,63 ppm) tal como Cc (46,43-48,93 ppm), Cd, por sua vez fica ligeiramente mais, desblindado (53,20-57,91 ppm). Relativamente a Ce, sinal característico de sais de guanidínio, aparece praticamente inalterado na conversão de 10a em 11a, aparecendo neste último ligeiramente mais desblindado. Tanto no espectro de ^1H -RMN como no de ^{13}C -RMN em vez de aparecer um sinal a corresponder a um tipo de carbono ou hidrogénio, aparecem múltiplos sinais, esse facto deve-se principalmente ao volume dos múltiplos átomos de bromo presentes em 11a, nesse contexto as diversas ligações do catião não vão rodar livremente podendo haver uma ou várias conformações

preferenciais podendo apresentar-se o catião de forma assimétrica, neste contexto quatro grupos substituintes (contendo sinais tipo **b**, **c** e **d**) à volta do catião guanidínio poderão não ser equivalentes, tal como os dois grupos substituintes correspondendo aos sinais do tipo **a**, resultando numa multiplicidade de sinais para cada tipo de desvio. De qualquer forma como o objectivo principal era a obtenção de novos sais de guanidínio densos e não o estudo da conformação ou conformações preferenciais em compostos como **11a**, assim este último tópico não foi aprofundado. O espectro de ^{13}C -RMN ainda evidencia que a reacção de bromação foi completa já que houve desaparecimento total de sinais na gama 121-131 ppm, ou em zonas próximas, característicos de carbonos olefinicos. Relativamente ao espectro de infravermelho este indica a presença de todos os tipos de ligações/funcionalidades presentes no catião pretendido, tal como a presença de ligações C-H alifáticas (2935 cm^{-1}), ligações C-N de guanidínio (1585 e 1517 cm^{-1}) e ligações C-Br (655 e 575 cm^{-1}). Neste espectro fica ainda evidente a presença de anião PF_6^- (840 cm^{-1}). O espectro de massa (FAB+) obtido confirma a presença do catião pretendido com sinais na gama $m/z = 881$ - 894 (887 combinação de isótopos maioritária) a representar precisamente o catião, neste espectro aparecem ainda sinais (combinação de isótopos maioritária) a $m/z = 807$, 727 , 645 e 568 indicando a perda de 1, 2, 3 e 4 átomos de bromos nas fragmentações realizadas para a obtenção de espectro de massa, este perfil confirma a presença do catião totalmente bromado na estrutura **11a**. Ficou assim demonstrado que esta via sintética é válida para a obtenção de sais de guanidínio bromados puros com potencial para serem novos materiais de elevada densidade. Ainda assim tanto o composto **10a** como **11a** foram obtidos como sólidos à temperatura ambiente quando o ideal seria a obtenção de líquidos que pudessem ter elevada densidade para eventualmente poderem ser utilizados como meio reaccional à temperatura ambiente ou como material inercial em instrumentos de navegação, **11a** não apresenta ponto de fusão inferior à sua temperatura de decomposição ($137,5^\circ\text{C}$). Adicionalmente **11a** foi obtido com um rendimento muito baixo (13,8%). Na tentativa de melhorar estes resultados, numa primeira abordagem procurou-se usar um outro anião em vez do hexafluorofosfato, desta forma usou-se um esquema de síntese modificado relativamente ao **Esquema II.24** (**Esquema II.25**):



Esquema II.25: Método C - Método de síntese modificado em relação a Esquema 24 – neste caso é usado o anião bis(trifluorometanosulfonil)imidato em vez de hexafluorofosfato.

Relativamente ao método de síntese representado no **Esquema II.25 – Método C**, ele é praticamente idêntico ao **Esquema II.24 – Método B**, sendo usado o mesmo material de partida, o cloreto **6a**. Este composto foi sujeito a reacção de troca de anião na presença de bis(trifluorometanosulfonil)imidato de lítio em vez de hexafluorofosfato de potássio (**Esquema II.24**) sendo obtido composto **12a** e posteriormente **13a** resultado de reacção de bromação realizada em condições semelhantes às representadas no **Esquema II.24**. Os compostos **12a** e **13a** possuem como anião o bis(trifluorometanosulfonil)imidato, este anião normalmente aparece associado a sais de baixo ponto de fusão ou líquidos à temperatura ambiente. Como resultado do primeiro passo de síntese (reacção de troca de anião) foi isolado um líquido amarelo que foi caracterizado por espectroscopia de ^1H -RMN, ^{13}C -RMN, Infravermelho e análise elementar (**Tabela II.34**):

Tabela II.34: Dados reaccionais, espectrais e análise elementar do composto intermediário 12a à síntese de sal de guanidínio de elevada densidade 13a.

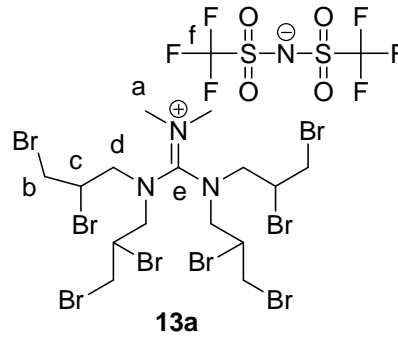
	Tempo de reacção η (%)	^1H -RMN δ (ppm) J (Hz)	^{13}C -RMN δ (ppm)	IV(filme ou KBr) $\nu_{\text{máx}}$ (cm^{-1})	Análise elementar
<p>12a</p>	56 horas 86,8%	2,93 (6H, s) Ha 3,74 (8H, sl) Hd 5,31-5,35 (8H, m) Hb 5,68 (4H, m) Hc	40,27 Ca 52,39, 52,86 Cd 114,87, 118,06 121,25, 124,44 Cf 121,75, 121,97 Cb 130,44 Cc 162,95 Ce	IV (filme) 3087 (CH olefínica) 2985 (CH alifática) 2925 (CH alifática) 1645 (C-N) 1541 (C-N) 1527 (C-N) 1348 ((CF_3SO_2) $_2\text{N}^-$) 1184 ((CF_3SO_2) $_2\text{N}^-$) 993 (CH olefínica) 935 (CH olefínica) 877 (CH olefínica) 865 (CH olefínica)	composição calculada para $\text{C}_{17}\text{H}_{26}\text{N}_4\text{F}_6\text{O}_4\text{S}_2 \cdot 2,9\text{H}_2\text{O}$ (%): N: 9,65 C: 35,16, H 5,52. composição determinada (%): N 9,87, C 34,67, H 5,00.

η – rendimento (%), δ - desvio químico, J – constante de acoplamento, $\nu_{\text{máx}}$ – energia representada em número de onda.

Tendo em conta os resultados apresentados na **Tabela II.34** pode-se observar a presença do catião pretendido já que no espectro de ^1H -RMN aparecem todos os sinais com as respectivas áreas a corresponder aos tipos/número de protões presentes no catião da estrutura **12a**. A destacar neste espectro o sinal correspondente aos protões do tipo **Ha** a 2,93 ppm e o sinal a 3,74 ppm correspondendo a protões do tipo **Hd**. O espectro de ^{13}C -RMN confirma a presença do catião pretendido podendo-se destacar o sinal do carbono quaternário **Ce** a 162,95 ppm característico a catiões de guanidínio. O espectro de ^{13}C -RMN ainda fornece informação relativamente à presença do anião pretendido o bis(trifluorometanosulfonil)imidato aparecendo quatro sinais a 114,87, 118,06, 121,25 e 124,44 ppm referentes ao carbono **Cf** resultado do acoplamento deste carbono com os três átomos de flúor a que está ligado. Relativamente ao espectro de infravermelho, este apresenta bandas correspondentes a todos os tipos de ligação/funcionalidades presentes tanto no catião como no anião. Neste caso pode-se destacar a banda a 3087 cm^{-1} , característica de ligações C-H olefinicas, as bandas a 2985 e 2925 cm^{-1} características de ligações C-H alifáticas e as bandas a 1645 , 1541 e 1527 cm^{-1} a indicar a presença de ligações C-N de guanidínio, todas estas bandas referentes ao catião; relativamente ao anião aparecem bandas a 1348 e 1184 cm^{-1} , que não estavam presentes no espectro do cloreto **6a**, e que indicam a presença do novo anião o bis(trifluorometanosulfonil)imidato. Os dados da análise elementar indicam que o composto pretendido está puro muito embora contenha

quantidade bastante considerável de água. Este composto (**12a**) foi sujeito a reacção de bromação sendo obtido um sólido castanho que foi analisado espectroscopicamente (^1H -RMN, ^{13}C -RMN e infravermelho) e por análise elementar sendo obtidos os seguintes resultados (**Tabela II.35**):

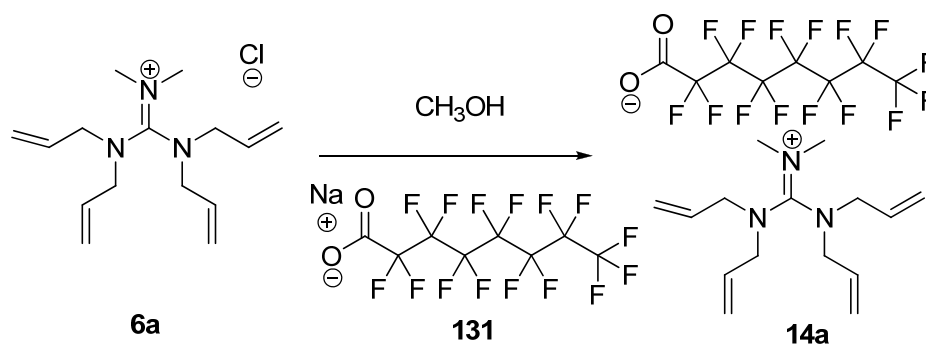
Tabela II.35: Dados reaccionais, espectrais e análise elementar de sal de guanidínio bromado 13a.

	Tempo de reacção η (%)	^1H -RMN δ (ppm) J (Hz)	^{13}C -RMN δ (ppm)	IV(filme ou KBr) $\nu_{\text{máx}}$ (cm^{-1})	Análise elementar
 <p>13a</p>	51 horas 44,3%	3,32-3,34 (8H, m) Hd 3,39-3,41 (6H, m) Ha 3,80-3,84 (8H, m) Hb 3,99-4,05 (4H, m) Hc	32,95, 33,08, 33,30, 33,67, 33,84, 34,27 Ca 41,94, 42,47, 42,86 Cb 45,41 e 45,55 Cc 53,71, 53,96, 54,29, 54,60, 57,11 Cd 114,99, 118,19, 121,38, 124,58 Cf 163,85, 164,64, 165,71 Ce	IV (KBr) 3032 (CH alifática) 3003 (CH alifática) 2957 (CH alifática) 2930 (CH alifática) 1582 (C-N) 1522, (C-N) 1344 ((CF ₃ SO ₂) ₂ N ⁻) 1190 ((CF ₃ SO ₂) ₂ N ⁻) 652 (C-Br) 569 (C-Br)	Composição calculada para C ₁₇ H ₂₆ Br ₈ F ₆ N ₄ O ₄ S ₂ (%): N: 4,80 C: 17,48, H 2,24. composição determinada (%): N 4,40, C 17,04, H 2,49.

η – rendimento (%), δ - desvio químico, J – constante de acoplamento, $\nu_{\text{máx}}$ – energia representada em número de onda.

Através da análise à **Tabela II.35** pode-se observar a presença do catião já que nos dados do espectro ^1H -RMN há correspondência entre os sinais deste espectro e os protões do catião do composto **13a**. Comparando os espectros do precursor **12a** e do produto bromado **13a** pode-se observar no espectro deste último a desblindagem dos protões Ha, por sua vez, os protões Hb, Hc e Hd aparecem mais blindados. Relativamente ao espectro de ^{13}C -RMN, este confirma a presença do catião pretendido havendo correspondência entre os sinais deste espectro e os tipos de carbono presentes na estrutura **13a**, relativamente ao espectro de ^{13}C -RMN de **12a**, o espectro de **13a** indica a blindagem de Ca, Cb e Cc, por sua vez Cd aparece mais desblindado tal como Ce. Este espectro indica ainda a presença do anião pretendido aparecendo quatro sinais a 114,99, 118,19, 121,38 124,58 ppm resultado do acoplamento do carbono Cf com três átomos de flúor a que está ligado. Tendo em conta a ausência de sinais por volta de 130 ppm característicos de carbonos olefinicos pode-se considerar que a reacção de bromação para a obtenção de **13a** foi completa. Comparando os espectros

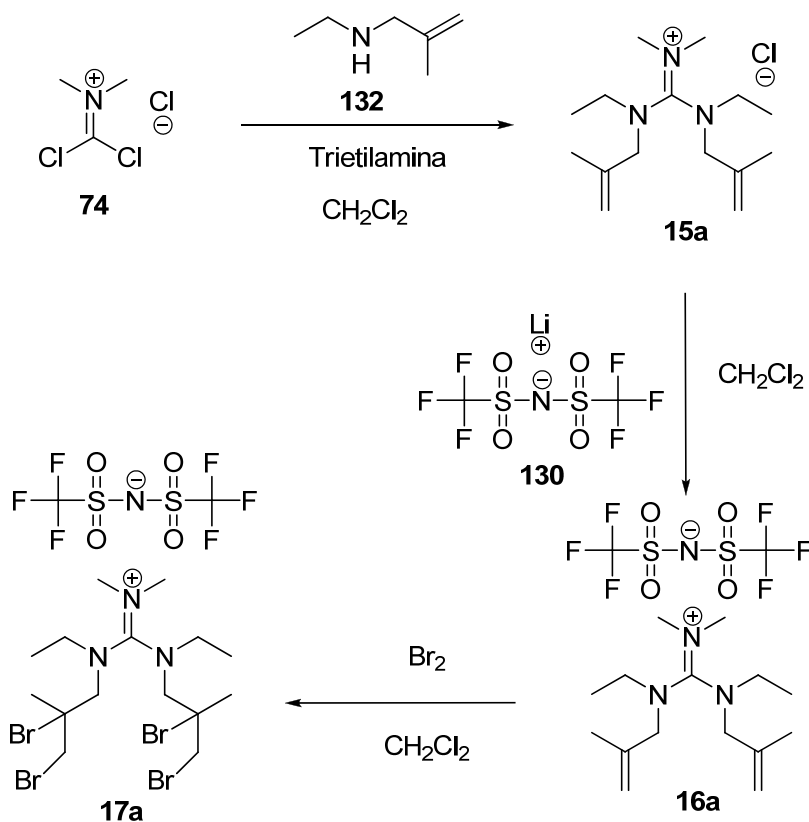
de ^1H -RMN e ^{13}C -RMN de **12a** e **13a** pode-se observar no caso deste último uma multiplicidade de sinais a corresponder a cada tipo de carbono e hidrogénio, esta multiplicidade de sinais está relacionada com o volume ocupado pelos oito átomos de bromo presentes no catião de **13a** fazendo com que a rotação das ligações não seja livre conduzindo a uma ou várias conformações preferenciais podendo estar quebrada a simetria associada aos quatro substituintes correspondendo a sinais do tipo **b**, **c** e **d**, tal como a simetria associada aos dois grupos substituintes correspondendo a sinais do tipo **a**, não sendo grupos substituintes equivalentes tanto ao nível conformacional como ao nível espectroscópico (^1H -RMN e ^{13}C -RMN). Analisando o espectro de Infravermelho pode-se observar correspondência entre as bandas obtidas e ligações/grupos funcionais presentes em **13a**. A destacar as bandas na gama $3032\text{--}2930\text{ cm}^{-1}$ típicas de ligações C-H alifáticas, as bandas a 1582 e 1522 cm^{-1} características de ligações C-N de catião guanidínio e as bandas a 652 e 569 cm^{-1} características de ligações C-Br. Este espectro indica ainda a presença do anião pretendido (1344 e 1190 cm^{-1}). O resultado da análise elementar evidência a alta pureza do composto obtido já que há correspondência entre a composição determinada e a composição calculada para a estrutura **13a**. Na síntese do composto **13a** o rendimento reaccional foi de 44,3% sendo usado 8 equivalentes de Br_2 e 51 horas de tempo reaccional. A reacção foi seguida por TLC começando aparecer a evidência de um único produto a partir das 26 horas reaccionais, ainda assim a reacção prosseguiu durante mais 25 horas de forma assegurar máximo rendimento. Apesar de ter sido obtido sal de guanidínio bromado com um rendimento mais aceitável (44,3% para **13a** vs 13,8% de sal **11a**) o composto **13a** obtido é sólido à temperatura ambiente não apresentando ponto de fusão inferior à sua temperatura de decomposição (143°C). Tendo em conta que o intermediário **12a** se apresenta líquido à temperatura ambiente (ainda que evidencie a presença de água) e o produto da bromação **13a** se apresenta sólido à temperatura ambiente com ponto de fusão superior a 143°C , pôs-se a hipótese de o excesso número de bromos (8) estar a induzir um elevado ponto de fusão. Tendo em conta essa hipótese procurou seguir-se duas vias: 1) Partir do precursor cloreto **6a** e por reacção de troca de anião introduzir anião que induzisse a formação de sal de elevada densidade sem ser necessário realizar reacção de bromação (**Esquema II.26**) e 2) Sintetizar um catião diferente que não apresentasse um número de bromos tão elevado (**Esquema II.27**).



Esquema II.26: Método D - Método de síntese baseado na troca de cloreto por anião perfluorooctanoato, este tipo de estrutura organo-fluorada aparece associada a compostos de elevada densidade.

No método de síntese representado no **Esquema II.26 – Método D**, à semelhança dos métodos de síntese representados nos **Esquemas II.24 (Método B)** e **II.25 (Método C)**, é usado como ponto de partida o cloreto de *N,N*-dimetil-*N',N',N'',N''*-tetraalilguanidínio **6a**, mas neste caso, em vez de ser introduzido o hexafluorofosfato ou o bis(trifluorometanosulfonil)imidato como aniões, é introduzido o anião perfluorooctanoato, anião contendo múltiplos átomos de flúor (15 átomos de flúor). Foi introduzido este anião pois havia referência na literatura a líquidos iônicos de elevada densidade contendo múltiplos átomos de flúor no catião e/ou no anião.⁹¹ Neste contexto antevia-se que a introdução do perfluorooctanoato fosse uma boa escolha tendo em vista a obtenção de um novo líquido iônico denso.

Nesta fase foi ainda seguida outra via (**Esquema II.27 – Método E**) que consistia na preparação de composto **17a** contendo catião com número mais reduzido de átomos de bromo relativamente aos compostos bromados preparados até aqui (**11a** e **13a**).

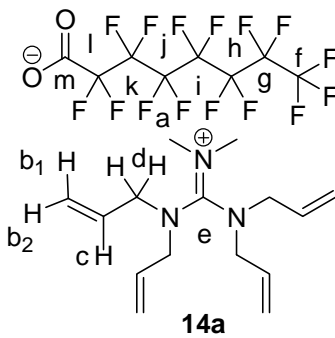


Esquema II.27: Método E - Método de síntese para a obtenção de sal de elevada densidade 17a que não contenha número elevado de átomos de bromos na estrutura do catião.

No método de síntese representado no **Esquema II.27 (Método E)** foi preparado sal de guanidínio **15a** a partir de cloreto de *N,N*-dimetilfosgenimínio (**74**) e amina secundária **132** contendo apenas um grupo olefinico. O composto **15a** obtido foi usado como ponto de partida para a obtenção de sal de bis(trifluorometanosulfonil)imidato (**16a**) intermediário para a obtenção do produto de bromação **17a**. A metodologia usada a partir do composto **15a** para a obtenção de material de elevada densidade **17a** é idêntica à representada no **Esquema II.25 (Método C)** com a diferença de se estar a obter um produto com quatro átomos de bromo em vez de oito.

Seguindo o **método D** representado no **Esquema II.26** foi obtido no final um composto líquido amarelo que foi caracterizado por espectroscopia ^1H -RMN, ^{13}C -RMN e infravermelho assim como por análise elementar (**Tabela II.36**).

Tabela II.36: Dados reaccionais, espectrais e análise elementar de sal de guanidínio 14a contendo anião perfluorooctanoato.

	Tempo de reacção η (%)	^1H -RMN δ (ppm) J (Hz)	^{13}C -RMN δ (ppm)	IV(filme ou KBr) $\nu_{\text{máx}}$ (cm^{-1})	Análise elementar
 <p>14a</p>	56 horas 87,6%	2,98 (6H, s) Ha 3,75 (8H, sl) Hd 5,27-5,31 (8H, m) Hb 5,70 ppm (4H, sl) Hc	40,39 Ca 52,40 e 52,80 Cd 110,83, 111,14 e outros sinais diminutos nesta zona espectral Cf-I 121,45, 122,61 Cb 130,87 Cc 160,06, 160,29, 160,52 Cm 163,06 Ce	IV (filme) 3084 (CH olefinica) 3015 (CH olefinica) 2984 (CH alifática) 2920 (CH alifática) 1643 (CO_2^-) 1589 (C-N) 1529 (C-N) 1420 (CO_2^-)	composição calculada para $\text{C}_{23}\text{H}_{26}\text{N}_3\text{O}_2\text{F}_{15}$ (%): N: 6,35 C: 41,76, H 3,96. composição determinada (%): N 6,49, C 41,24, H 4,16.

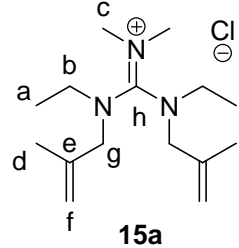
η – rendimento (%), δ - desvio químico, J – constante de acoplamento, $\nu_{\text{máx}}$ – energia representada em número de onda.

Considerando os resultados apresentados na **Tabela II.36** e olhando concretamente para os dados espectroscópicos de ^1H -RMN é possível confirmar a presença do catião pretendido já que há correspondência entre os sinais apresentados e os prótons do catião pertencente à estrutura **14a**. Relativamente ao espectro de ^{13}C -RMN este confirma a presença do catião pretendido pela correspondência entre sinais apresentados e tipos de carbonos presentes no catião pretendido, neste caso destaca-se o sinal a 163,06 ppm correspondente ao carbono Ce, característico de catião guanidínio. Este espectro evidencia a presença do anião pretendido já que aparecem um série de sinais perto de 110 ppm correspondentes a carbonos Cf-I da cadeia fluorada do anião, adicionalmente aparece sinal a 160,52 ppm pertencente ao carbono Cm do grupo carboxilato contido no anião. Precisamente o grupo carboxilato contido no anião aparece evidente no espectro de Infravermelho aparecendo bandas a 1643 e 1420 cm^{-1} características desse grupo. Este espectro confirma ainda a presença do catião pretendido aparecendo bandas a 3084 e 3015 cm^{-1} características de ligações CH olefinicas, adicionalmente, aparecem no espectro bandas a 2984 e 2920 cm^{-1} dando conta da presença de ligações CH alifáticas, tal como bandas a 1589 e 1529 cm^{-1} correspondentes a ligações CN de guanidínio. A identificação e a pureza do composto pretendido são confirmadas por análise elementar havendo correspondência entre a composição determinada experimentalmente e a composição calculada para a combinação pretendida de catião e anião. Este composto **14a** é líquido à temperatura

ambiente, muito embora há que olhar com cautela para o estado físico deste composto dada a quantidade de água titulada por Karl Fischer (3,1% em massa)

Relativamente ao **Método E** de síntese apresentado no **Esquema II.27**, como resultado do primeiro passo de síntese foi obtido um sólido branco que aparece aqui caracterizado na **Tabela II.37**.

Tabela II.37: Dados reaccionais, espectrais e análise elementar de sal de guanidínio 15a.

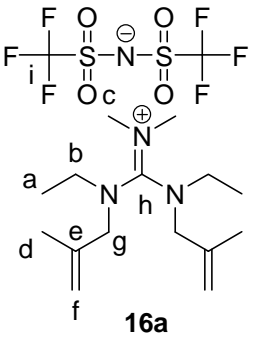
	Tempo de reacção η (%)	^1H -RMN δ (ppm) J (Hz)	^{13}C -RMN δ (ppm)	IV(filme ou KBr) $\nu_{\text{máx}}$ (cm^{-1})	Análise elementar	Espectro de massa exacta catião (IE)
	69 horas	1,16 (6H, m) Ha	12,96, 13,25 Ca	IV (KBr)	composição calculada para $\text{C}_{15}\text{H}_{30}\text{N}_3\text{Cl}$ (%): N: 14,60 C: 62,58, H 10,50. composição determinada (%): N 14,48, C 62,01, H 10,03.	Experimental 252,244117
	44,6%	1,66 (6H, s) Hd 3,07-3,18 (10H, complexo) Hb,c 3,81-3,91 (4H, m) Hg 5,02-5,05 (4H, m) Hf	20,32 Cd 40,72 e 41,00 Ce 44,41 e 44,72 Cb 55,54, 56,04 Cg 117,82 Cf 137,23, 137,65 Ce 163,45 Ch	3083 (CH olefínica) 2976 (CH alifática) 2939 (CH alifática) 1650 (C-N) 1579 (C-N) 1537 (C-N) 989 (CH olefínica) 918 (CH olefínica) 862 (CH olefínica)		Calculado ($\text{C}_{15}\text{H}_{30}\text{N}_3^+$ requer 252,243973)

η – rendimento (%), δ - desvio químico, J – constante de acoplamento, $\nu_{\text{máx}}$ – energia representada em número de onda, IE – impacto electrónico.

De acordo com os dados da **Tabela II.37** foi obtido o catião pretendido presente na estrutura **15a**, já que pelos dados do espectro de ^1H -RMN há correspondência entre os sinais/áreas obtidos e os tipos/número de protões presentes no catião da estrutura **15a** aparecendo sinal a 1,16 ppm correspondendo aos 6 protões do tipo Ha, por sua vez na gama 3,07-3,18 ppm aparece multiplete com área correspondente a dez protões, quatro protões Hb e seis protões Hc. Tendo em conta a área correspondente a Hc e a área de Ha pode-se afirmar que a reacção de obtenção de catião guanidínio, a partir de amina secundária e cloreto de *N,N*-dimetilfosgenimínio, foi completa tendo entrado 2 moléculas de amina secundária a substituir 2 átomos de cloro contidos na molécula de cloreto de *N,N*-dimetilfosgenimínio (**Esquema II.27**). Continuando a análise ao espectro de ^1H -RMN, foi possível identificar singuleto a 1,66 ppm com área correspondente a seis protões Hd, foi ainda observado sinal na gama 3,81-3,91 ppm correspondendo a quatro protões Hg em adição ao sinal apresentado na gama 5,02-5,05 ppm correspondendo a quatro protões Hf. Relativamente ao espectro de ^{13}C -RMN este confirma a presença do catião pretendido, havendo correspondência entre os sinais apresentados e os

carbonos do catião da estrutura **15a**, concretamente, aparecem sinais a 12,96 e 13,25 ppm correspondentes a carbonos **Ca**, seguidamente aparece sinal a 20,32 ppm correspondente a carbonos do tipo **Cd**. Significativamente mais desblindados aparecem carbonos do tipo **Cc** correspondendo a sinais a 40,72 e 41,00 ppm. Por sua vez a 44,41 e 44,72 ppm aparecem sinais atribuídos a carbonos do tipo **Cb**, seguidos por carbonos **Cg** correspondendo a sinais a 55,54 e 56,04 ppm. Significativamente mais desblindados aparecem os carbonos **Cf** e **Ce** olefinicos correspondendo respectivamente a sinal a 117,82 ppm e a dois sinais a 137,23 e 137,65 ppm. Finalmente aparece a 163,45 ppm o sinal mais característico de catiões de guanidínio, correspondendo concretamente ao carbono **Ch**. Neste espectro nota-se que determinados carbonos vão corresponder a mais que um sinal, este facto deve-se muito provavelmente a constrangimentos na rotação de algumas ligações químicas já que o catião é olefinico e ramificado. Relativamente ao espectro de Infravermelho este indica a presença de todas as funcionalidades/tipos de ligação presentes no catião pretendido, sendo identificada banda a 3083 cm^{-1} correspondente a ligações CH olefinicas, adicionalmente aparecem no espectro bandas a 2976 e 2939 cm^{-1} correspondentes a ligações CH alifáticas e bandas a 1650, 1579 e 1537 correspondentes a ligações CN de guanidínio. Se havia alguma dúvida relativamente à identidade do catião obtido esta é completamente dissipada pelos dados de espectro de massa exacta sendo detectado fragmento cuja massa detectada corresponde à massa calculada para o catião pretendido, esta correspondência ocorre até à segunda casa decimal inclusive. Os dados da análise elementar confirmam a identidade e pureza do sólido branco obtido havendo correspondência entre composição determinada experimentalmente e a composição calculada para a estrutura **15a**. A reacção para obtenção deste sal de guanidínio durou 69 horas e mesmo assim o rendimento foi de 44,6%. Continuando a percorrer o **Esquema II.27 (Método E)** partindo de **15a** é realizada reacção de troca de anião sendo obtido no final um líquido amarelo que foi caracterizado (**Tabela II.38**):

Tabela II.38: Dados reaccionais, espectrais e análise elementar de sal de bis(trifluorometanosulfonil)imidato 16a.

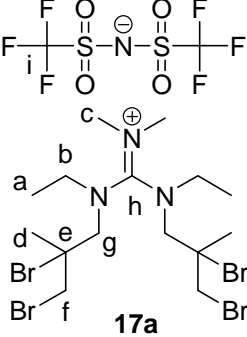
	Tempo de reacção η (%)	^1H -RMN δ (ppm) J (Hz)	^{13}C -RMN δ (ppm)	IV(filme ou KBr) $\nu_{\text{máx}}$ (cm^{-1})	Análise elementar
 <p>16a</p>	56 horas	1,15-1,20 (6H, m) Ha	12,21, 12,59 Ca	IV(filme)	composição calculada para $\text{C}_{17}\text{H}_{30}\text{N}_4\text{O}_4\text{S}_2$ F₆ (%) : N: 10,52 C: 38,34, H 5,68. composição determinada (%) : N 10,51, C 38,16, H 5,56.
	70%	1,70 (6H, s) Hd	19,15, 19,65 Cd	3082 (CH olefinica)	
		2,94-3,03 (6H, m) Hc	39,76, 40,01, 40,22 Cc	2978 (CH alifática)	
		3,20-3,28 (4H, m) Hb	43,78 e 44,39 Cb	2941 (CH alifática)	
		3,69-3,79 (4H, m) Hg	55,08, 55,66 Cg	1655 (C-N)	
		5,07-5,11 ppm (4H, m) Hf	114,60, 117,55, 120,99, 124,19 Ci	1570 (C-N)	
			117,80, 117,99 Cf	1527 (C-N)	
			136,72, 136,99 Ce	1350 ((CF_3SO_2) ₂ N ⁺)	
			162,74 e 163,15 Ch	1134 ((CF_3SO_2) ₂ N ⁺)	
				987 (CH olefinica)	
				914 (CH olefinica)	
				870 (CH olefinica)	
				820 (CH olefinica)	

η – rendimento (%), δ - desvio químico, J – constante de acoplamento, $\nu_{\text{máx}}$ – energia representada em número de onda.

Os dados apresentados na **Tabela II.38** confirmam a presença do catião pretendido (contido na estrutura **16a**) em especial os dados do espectro de ^1H -RMN, havendo correspondência entre localização/área dos sinais obtidos e o tipo e número de prótons presentes no catião pretendido. Relativamente ao espectro de ^{13}C -RMN este confirma a presença do catião pretendido havendo correspondência entre os sinais obtidos e os tipos de carbonos presentes no catião, neste caso há a destacar os sinais a 162,74 e 163,15 ppm característicos de carbono **Ch** de núcleo guanidínio. Este espectro também apresenta sinais correspondentes ao anião pretendido, mais precisamente quatro sinais a 114,60, 117,55, 120,99 e 124,19 ppm correspondentes ao carbono **Ci** do anião que aparece ligado a três átomos de flúor. No que diz respeito ao espectro de Infravermelho este indica a presença de todos os tipos de ligações/funcionalidades presentes tanto no catião como no anião, relativamente àquele aparece banda a 3082 cm^{-1} correspondendo a ligações do tipo CH olefinicas, adicionalmente aparecem bandas a 2978 e 2941 cm^{-1} correspondentes a ligações CH alifáticas, aparecem ainda neste espectro bandas a 1655 , 1570 e 1527 cm^{-1} correspondentes a ligações CN de catião guanidínio. Relativamente ao anião aparecem duas banda característica a 1350 cm^{-1} correspondendo ao anião

bis(trifluorometanosulfonil)imidato. Finalmente a identidade e pureza do composto pretendido são confirmadas por análise elementar, havendo correspondência entre a composição determinada experimentalmente e a composição calculada para o sal **16a** contendo catião e anião específicos na proporção 1:1. A partir do composto **16a** obtido foi realizada reacção de bromação de acordo com o terceiro passo reaccional do **Esquema II.27 (Método E)** sendo obtido no final um sólido amarelo que foi caracterizado, sendo apresentados os respectivos dados da caracterização na **Tabela II.39**.

Tabela II.39: Dados reaccionais, espectrais e análise elementar de sal tetrabromado 17a.

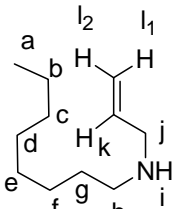
	Tempo de reacção η (%)	^1H -RMN δ (ppm) J (Hz)	^{13}C -RMN δ (ppm)	IV(filme ou KBr) $\nu_{\text{máx}}$ (cm^{-1})	Análise elementar	Espectro de massa exacta catião (IE)
	56 horas	1,33-1,43 (6H, m) Ha	12,34-14,05 Ca	IV(KBr)	Composição calculada para $\text{C}_{17}\text{H}_{30}\text{N}_4\text{O}_4\text{S}_2\text{F}_6\text{Br}_4$ (%): N: 6,57 C: 23,96, H 3,55. composição determinada (%): N 6,57, C 23,93, H 3,85.	Experimental 571,9135
	75%	1,87-2,01 (6H, m) Hd 2,97-2,99 (4H, m) Hb 3,27-3,29 (6H, m) Hc 3,66-3,71 (4H, m) Hg 3,79-3,92 (4H, m) Hf	29,08, 29,26, 29,31 Cd 39,92 Cc 40,89, 41,60, 41,82 Cb 47,96, 48,11, 48,22, 48,59 Ce 54,20, 54,79, 54,97, 57,02, 57,24 Cf 65,01, 66,79, 66,99, 67,27 Cg 115,06, 118,25, 121,45, 124,65 Ci 166,81, 167,42 Ch	2995 (CH alifática) 2977 (CH alifática) 2945 (CH alifática) 1589 (C-N) 1577 (C-N) 1348 ((CF_3SO_2) $_2\text{N}^+$) 635 (C-Br)		Calculado ($\text{C}_{15}\text{H}_{30}\text{Br}_4\text{N}_3^+$ requer 571,9132).

η – rendimento (%), δ - desvio químico, J – constante de acoplamento, $\nu_{\text{máx}}$ – energia representada em número de onda, IE – impacto electrónico.

Tendo em conta os dados apresentados na **Tabela II.39** confirma-se a presença do catião bromado pretendido, de acordo com estrutura **17a**, especialmente pelos dados do espectro de ^1H -RMN evidenciando correspondência entre a localização dos sinais/áreas obtidas e o tipo e número de prótons presentes no catião pretendido. Concretamente, na conversão de **16a** a **17a** os prótons Ha ficaram ligeiramente mais desblindados aparecendo agora na gama 1,33-1,43 ppm os sinais correspondentes, o mesmo a acontecer com os prótons Hd que ficaram agora ligeiramente mais desblindados correspondendo agora a sinais na gama 1,87-2,01 ppm. Os prótons Hb ficam ligeiramente mais blindados aparecendo agora sinais na gama 2,97-2,99 ppm, por sua vez os prótons

Hc ficam mais desblindados correspondendo a sinais na gama 3,27-3,29 ppm. Relativamente aos prótons Hg estes ficam praticamente inalterados correspondendo agora a sinais na gama 3,66-3,71 ppm. Finalmente os prótons Hf ficam agora significativamente mais blindados aparecendo agora sinais na gama 3,79-3,92 ppm, os prótons Hf no composto **16a** estavam ligados a carbono olefinico enquanto que agora estão ligados a carbono não olefinico que está ligado a átomo de bromo. Relativamente ao espectro de ^{13}C -RMN este confirma a presença do catião pretendido havendo correspondência entre os sinais do espectro e os tipos de carbono presentes no catião alvo. A destacar neste espectro os sinais correspondentes ao carbono Ch de núcleo guanidínio correspondendo a sinais 166,81 e 167,42 ppm. A destacar ainda neste espectro o desaparecimento de sinais a 136 ppm correspondente a carbono olefinico Ce na estrutura **16a**, agora na estrutura **17a** o carbono Ce aparece agora representado por sinais a 47,96, 48,11, 48,22 e 48,59 ppm, estes dados dão a indicação que a reacção de bromação foi completa. Relativamente ao anião foram detectados neste espectro quatro sinais a 115,06, 118,25, 121,45 e 124,65 ppm característicos de carbono Ci do anião. Tanto no espectro de ^1H -RMN como no de ^{13}C -RMN cada tipo de próton/carbono aparece representado por uma multiplicidade de sinais, esse facto está relacionado com constrangimentos na rotação das ligações devido ao elevado volume dos quatro átomos de bromo introduzidos. Relativamente ao espectro de infravermelho este evidencia a presença de todos os tipos de funcionalidades/tipos de ligação presentes tanto no catião como no anião, relativamente àquele aparecem bandas a 2995, 2977 e 2945 cm^{-1} correspondendo a ligações CH alifáticas, adicionalmente aparecem bandas a 1589 e 1577 cm^{-1} características de ligações CN de núcleo guanidínio, ainda relativamente ao catião aparece banda a 635 cm^{-1} característica de ligações CBr. No que diz respeito ao anião este é identificado por bandas a 1348 e 1174 cm^{-1} associadas a anião bis(trifluorometanosulfonil)imidato. A presença do catião bromado pretendido é confirmada por espectro de massa exacta sendo detectado fragmento cuja massa detectada corresponde à massa calculada para o catião pretendido, esta correspondência ocorre até à terceira casa decimal inclusive. Relativamente aos dados de análise elementar estes indicam a identidade e a pureza do composto **17a** pretendido havendo correspondência entre a composição determinada experimentalmente e a composição calculada para a combinação de catião e anião presentes no composto **17a**. Apesar do sucesso na síntese do composto **17a** este apresenta-se sólido à temperatura ambiente. Tendo em conta esse facto, procurou seguir-se outra estratégia baseada no conhecimento adquirido no estudo **II-3.4 d)**, nesse estudo ficou evidente que sais de guanidínio que apresentassem grupos substituintes assimétricos relativamente ao tamanho das cadeias tinham a tendência a apresentar baixo ponto de fusão. Procurou-se assim sintetizar um novo sal de guanidínio com estas características que depois pudesse ser bromado (**Esquema II.28**):

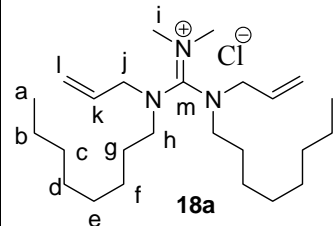
Tabela II.40: Dados espectroscópicos referentes à amina secundária 18c.

	Tempo de reacção η (%)	^1H -RMN δ (ppm) J (Hz)	^{13}C -RMN δ (ppm)
 <p>18c</p>	36 horas	0,86-0,89 (3H, t, J = 6,8) Ha 1,27-1,29 (10H, complexo) Hb-f 1,48-1,53 (2H, m) Hg 2,29 (1H, sl) Hi 2,60-2,64 (2H, t, J = 7,2) Hh 3,27-3,28 (2H, sl) Hj 5,11 (1H, d, J = 10,4) Hl ₂ 5,19 (1H, d, J = 17,2) Hl ₁ 5,88-5,98 (1H, m) Hk	14,03 Ca 22,60 Cb 27,31, 29,21, 29,46, 29,79, 31,78 Cc-g 49,23 Ch 52,28 Cj 116,20 Cl 136,34 Ck
	43,7%		

203

espectro é possível observar correspondência entre os restantes sinais e os respectivos protões assinalados na estrutura **18c**. O espectro de ^{13}C -RMN confirma a presença da amina secundária pretendida. A reacção de alquilação da amina durou 36 horas sendo obtido no final um rendimento de 43,7%. A amina secundária **18c** foi posteriormente usada na síntese do cloreto de guanidínio **18a** no segundo passo reaccional representado no **Esquema II.28**, como produto do segundo passo reaccional foi obtido um líquido laranja que foi caracterizado espectroscopicamente (**Tabela II.41**):

Tabela II.41: Dados espectroscópicos referentes a sal de guanidínio 18a contendo grupos substituintes diferenciados em termos de tamanho.

	Tempo de reacção η (%)	^1H -RMN δ (ppm) J (Hz)	^{13}C -RMN δ (ppm)	Espectro de massa exacta catião (IE)
	19 horas	0,58-0,67 (6H, m) Ha	13,52 Ca	Experimental 392,4005 Calculado ($\text{C}_{25}\text{H}_{50}\text{N}_3^+$ requer 392,3999)
	14%	0,97-1,05 (24H, complexo) Hb-g 2,90 (6H, s) Hi 3,01-3,03 (4H, m) Hh 3,53-3,69 (4H, m) Hj 5,13-5,24 (4H, m) Hl 5,49-5,67 (2H, m) Hk	22,04, 26,25, 26,46, 26,75, 27,20, 28,58, 30,63, 31,17 Cb-g 40,24 e 40,61 Ci 48,84, 49,27, 49,56, 49,74 Ch 52,18, 52,43, 52,70, 53,17 Cj 120,80, 121,06 Cl 130,39, 130,93 Ck 162,85 Cm	

η – rendimento (%), δ - desvio químico, J – constante de acoplamento, IE – impacto electrónico.

Tendo em conta os resultados apresentados na **Tabela II.41** foi possível isolar o sal de guanidínio **18a** pretendido. Concretamente, através da análise ao espectro de ^1H -RMN é possível afirmar que para o composto isolado a reacção foi completa havendo correspondência entre áreas dos sinais referentes aos protões Ha e Hi, a 0,58-0,67 (multiplete) e 2,90 ppm (singuleto) respectivamente. Há correspondência entre os restantes sinais deste espectro e os tipos de protões presentes no catião da estrutura alvo **18a**. Relativamente ao espectro de ^{13}C -RMN este confirma a presença do catião pretendido, destacando-se neste espectro o sinal a 162,85 ppm característico de

sais de guanidínio correspondendo especificamente ao carbono **Cm**. A presença do catião pretendido é confirmada inequivocamente através de espectro de massa exacta sendo identificado fragmento cuja massa detectada corresponde à massa calculada para o catião pretendido, a discrepância entre os dois valores é de 0,6 milésimos. O cloreto de guanidínio **18a** foi obtido como líquido à temperatura ambiente enquanto que os cloretos de guanidínio preparados para usar segundo os métodos **B**, **C**, **D** e **E** são sólidos à temperatura ambiente, este resultado coloca a hipótese de ser possível preparar um sal de elevada densidade líquido à temperatura ambiente baseado na unidade guanidínio. Apesar de ter sido obtido como líquido à temperatura ambiente o sal **18a** foi obtido com um rendimento muito baixo (14%) mesmo após reacção de 19 horas, esse baixo rendimento deve-se em parte à formação de produto secundário **18b** (**Figura II.34**):

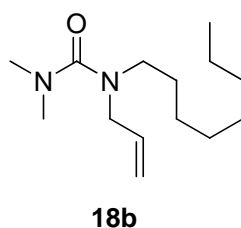
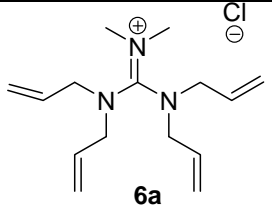
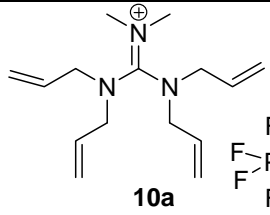
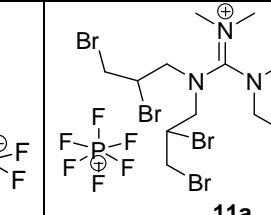


Figura II.34: 1-Alil-3,3-dimetil-1-octilureia – produto secundário 18b.

Segundo o **Esquema II.19** a ureia **18b** é produto de reacção de substituição de cloreto por uma molécula de amina secundária, o segundo cloreto não é substituído por uma molécula de amina secundária mas sim por uma molécula de água resultando na ureia representada na **Figura II.34**. Desta forma fica explicado o baixo rendimento em sal de guanidínio **18a** obtido, tendo em conta esse facto não foi possível isolar **18a** em quantidade suficiente para prosseguir com a sequência sintética de trocar anião cloreto por bis(trifluorometanosulfonil)imidato e posterior reacção de bromação de olefinas para obtenção de líquido iónico de elevada densidade.

Como resultado da execução dos métodos **B**, **C**, **D**, **E** e **F** foram isolados uma série de compostos que foram caracterizados relativamente ao ponto de fusão, densidade e presença de água. Relativamente ao **método B** intervieram três compostos, o cloreto **6a**, como precursor, o sal de hexafluorofosfato **10a** como intermediário e finalmente o produto final, o sal de hexafluorofosfato de guanidínio bromado **11a** (**Tabela II.42**).

Tabela II.42: Comparação ao nível do ponto de fusão e densidade entre compostos intervenientes no método B de síntese.

Composto			
Ponto de fusão (transição vítrea) (°C)	79,2	56,1	137,5 (temperatura de decomposição)
Densidade (g/cm ³)	1,06 ^c	1,21 ^c	1,55 ^c

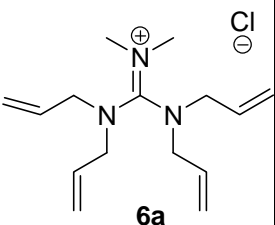
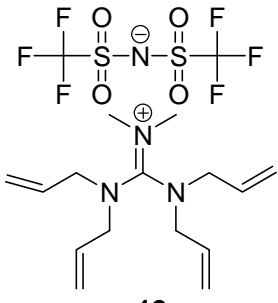
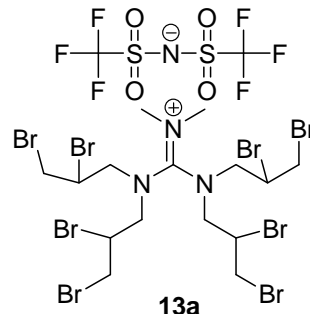
^c – densidade de sólidos foi medida por método indirecto descrito na Parte Experimental a 25°C.

Tendo em conta os resultados apresentados na **Tabela II.42** para os três compostos intervenientes no **método B** de síntese pode-se afirmar, comparando o precursor cloreto **6a** com o intermediário baseado em anião hexafluorofosfato **10a**, que a introdução do anião hexafluorofosfato permite a obtenção de densidades mais elevadas em relação ao anião cloreto (1,21 para sal de hexafluorofosfato vs 1,06 g/cm³ para cloreto). Observou-se também redução no ponto de fusão ao trocar o anião cloreto em **6a** por anião hexafluorofosfato em **10a**, em ambos os compostos o catião é o mesmo ficando assim evidente o efeito do anião. Ainda assim tendo em conta a presença de água na amostra contendo o cloreto **6a** há que olhar com a devida precaução para as propriedades físico-químicas associadas a esta amostra. No segundo passo de síntese usando o **método B** é obtido o sal de guanidínio bromado **11a** a partir do sal de hexafluorofosfato **10a**; em ambos os casos o anião é o mesmo, o hexafluorofosfato, neste caso vai estar em evidência nesta comparação diferenças entre os catiões de **10a** e **11a**, ou seja, vai ser comparado o efeito da conversão das posições olefinicas em **10a** para posições bromadas em **11a**. Após esta conversão pode-se observar um incremento significativo da densidade (1,55 em **11a** vs 1,21 g/cm³ em **10a**), tal como incremento do ponto de fusão (>137,5 em **11a** vs 56,1°C em **10a**). O bromo para além de ser um elemento pesado (aumentando a densidade) é também um elemento muito polarizável e volumoso, esses factos podem significar incremento da importância de interacções de Van der Waals, tal como diminuição dos graus de liberdade conformacionais do catião (evidente em espectro de ¹³C-RMN) reduzindo entropia de fusão, ambos os factores podem contribuir para o aumento do ponto de fusão.

Por sua vez no **método C** de síntese, em tudo idêntico ao método B excepto no uso do anião bis(trifluorometanosulfonyl)imidato em vez do anião hexafluorofosfato, intervêm três compostos, o precursor cloreto **6a** (tal como no **método B**), o intermediário sal de

bis(trifluorometanosulfonil)imidato **12a** e o produto final, o sal de guanidínio bromado **13a** (Tabela II.43):

Tabela II.43: Comparação ao nível do ponto de fusão e densidade entre compostos intervenientes no método C de síntese.

Composto			
Ponto de fusão (transição vítrea) (°C)	79,2	(-57,94)	143,2 (temperatura de decomposição)
Densidade (g/cm ³)	1,06 ^c	1,342 ^a	2,15 ^c

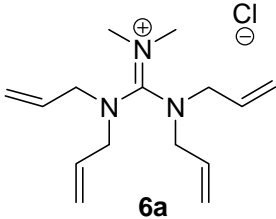
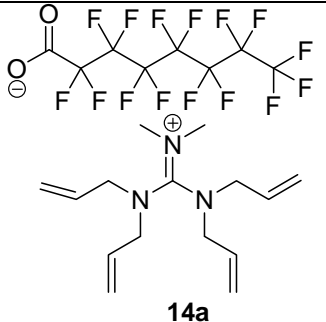
^a – densidade de líquidos em volumes superiores a 700 µl medida por densímetro Anton Paar DMA38 a 25°C. ^c – densidade de sólidos foi medida por método indirecto descrito na Parte Experimental a 25°C.

Tendo em conta os resultados apresentados na Tabela II.43, comparando o sal **6a** com o sal **12a**, saís que partilham o mesmo catião, pode-se ver claramente o efeito do anião, havendo um claro incremento na densidade (1,342 para **12a** vs 1,06 g/cm³ para **6a**) e uma diminuição significativa na temperatura de transição para o estado líquido na troca de cloreto em **6a** por anião bis(trifluorometanosulfonil)imidato em **12a**. O anião bis(trifluorometanosulfonil)imidato geralmente permite a obtenção de sais de baixo ponto de fusão já que para além de ter a carga negativa deslocalizada é um anião organo-fluorado, tipo de estrutura que apresenta interacções de baixa intensidade com outras estruturas orgânicas (neste caso o catião guanidínio), isto já que nas estruturas organo-fluoradas, devido a elevada electronegatividade do flúor, os carbonos associados apresentam carga parcial positiva algo que não acontece com as habituais estruturas orgânicas,¹¹⁰ estes factos somados permitem redução significativa do ponto de fusão. Relativamente à densidade o anião bis(trifluorometanosulfonil)imidato é um anião de dimensões mais próximas ao catião, provocando um aumento na densidade tal como foi observado no estudo II-3.3 c) da densidade de sais de imidazólio, por outro lado o anião bis(trifluorometanosulfonil)imidato é um anião fluorado, essa característica estrutural provoca em geral um aumento na densidade. No caso do sal de bis(trifluorometanosulfonil)imidato **12a** detectou-se presença de água (0,98% em massa). No segundo passo de síntese do método C vai ser obtido sal de guanidínio bromado **13a** a partir do sal de

bis(trifluorometanosulfonil)imidato **12a** anteriormente discutido. Tanto em **12a** como em **13a** o anião é o mesmo, o bis(trifluorometanosulfonil)imidato, diferindo no catião empregue, o catião de **12a** possui quatro posições olefinicas, enquanto que no catião em **13a** as posições olefinicas são substituídas por ligações carbono-bromo. Assim na conversão de **12a** em **13a** vai haver um incremento significativo da densidade (2,15 em **13a** vs 1,342 g/cm³ em **12a**) assim como um aumento muito significativo na temperatura de passagem para o estado líquido (>143,2 em **13a** vs -57,94°C em **12a**). Os motivos para estes incrementos muito significativos devem-se à massa do bromo, ao seu volume e polarizabilidade tal como foi discutido para o composto **11a** no âmbito do **método B** de síntese. Comparando o sal de hexafluorofosfato **10a** com o sal de bis(trifluorometanosulfonil)imidato **12a**, contendo ambos o mesmo catião, pode-se observar que o sal de bis(trifluorometanosulfonil)imidato permite a obtenção de densidade mais elevada (1,21 em **10a** vs 1,342 g/cm³ em **12a**) e temperatura de transição para o estado líquido mais baixa (56,1 em **10a** vs -57,94°C em **12a**). Comparando o efeito do anião ao nível de sais de guanidínio bromados, pode-se observar que no caso do sal de bis(trifluorometanosulfonil)imidato **13a** a densidade é muito mais elevada do que no caso do sal de hexafluorofosfato **11a** (2,15 em **13a** vs 1,55 g/cm³ em **11a**), neste caso a diferença de densidades é muito mais pronunciada do que entre **10a** e **12a**, em **13a**, as dimensões do anião são mais semelhantes às dimensões do catião do que no caso de **11a**.

Relativamente ao **método D** de síntese, que consiste na realização de um único passo reaccional, intervêm dois compostos, o reagente cloreto **6a** (à semelhança dos **métodos B** e **C** de síntese), e o produto reaccional o sal de perfluorooctanoato **14a** (Tabela II.44).

Tabela II.44: Comparação ao nível do ponto de fusão e densidade entre compostos intervenientes no método D de síntese.

Composto	 6a	 14a
Ponto de fusão (transição vítrea) (°C)	79,2	(-49,4)
Densidade (g/cm ³)	1,06 ^c	1,369 ^b

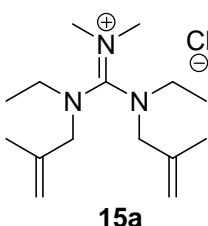
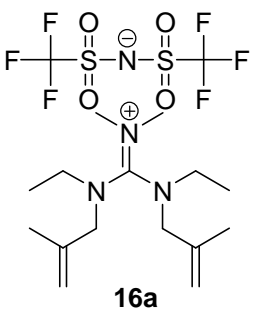
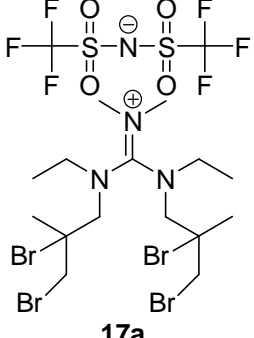
^b – densidade de líquidos em volumes inferiores a 700 µl medida em micropicnómetro a 21°C de acordo com método descrito na Parte experimental. ^c – densidade de sólidos foi medida por método indirecto descrito na Parte Experimental a 25°C.

Através da análise da **Tabela II.44** pode-se observar que a diferença entre o material de partida **6a** e o produto reaccional **14a** está no anião empregue, cloreto em **6a** e perfluorooctanoato em **14a**, partilhando ambos os compostos o mesmo catião, assim, observa-se claramente o efeito que o anião tem na densidade (e temperatura de transição para o estado líquido). Desta forma observa-se que o sal de perfluorooctanoato **14a** possui densidade significativamente mais elevada que o sal cloreto **6a** (1,369 vs 1,06 g/cm³ para **14a** e **6a** respectivamente). O anião perfluorooctanoato, à semelhança do anião bis(trifluorometanosulfonil)imidato, tem capacidade de deslocalizar a carga negativa sendo um anião organofluorado (ver discussão para sal de bis(trifluorometanosulfonil)imidato **10a** no âmbito da execução de **método C** de síntese), conduzindo significativamente ao incremento da densidade e diminuição da temperatura de transição para o estado líquido (-49,4 vs 79,2°C para **14a** e **6a** respectivamente, muito embora há que ter em conta a presença de 3,1% em massa de água no caso de **14a**). Comparando a densidade dos sais **6a**, **10a**, **12a** e **14a**, sais que partilham o mesmo catião diferindo no anião empregue, pode-se observar a seguinte relação relativamente à densidade:



Relativamente à execução do **método E**, foram preparados três sais de forma sequencial, ou seja, em primeiro lugar foi sintetizado o sal cloreto **15a**, que foi usado na síntese do sal de bis(trifluorometanosulfonil)imidato **16a** que por sua vez foi precursor na obtenção de sal de guanidínio bromado **17a** (**Tabela II.45**).

Tabela II.45: Comparação ao nível do ponto de fusão e densidade entre compostos intervenientes no método E de síntese.

Composto	 15a	 16a	 17a
Ponto de fusão (transição vítrea) (°C)	146,5	(-56,4)	74,3
Densidade (g/cm ³)	1,02 ^c	1,284 ^b	1,68 ^c

^b – densidade de líquidos em volumes inferiores a 700 µl medida em micropicnómetro a 22°C de acordo com método descrito na Parte experimental. ^c – densidade de sólidos foi medida por método indirecto descrito na Parte Experimental a 25°C.

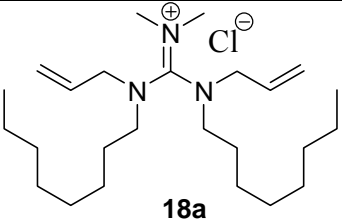
De acordo com a **Tabela II.45**, comparando os sais **15a** e **16a**, que partilham o mesmo catião diferindo no anião, pode-se observar um incremento na densidade do sal de bis(trifluorometanosulfonil)imidato **16a** em relação ao cloreto **15a** (1,284 vs 1,02 g/cm³ respectivamente), assim como uma diminuição significativa na temperatura de transição para o estado líquido (-56,4°C para sal de bis(trifluorometanosulfonil)imidato **16a** vs 146,5°C para cloreto **15a**), para os factos observados contribuem capacidade de deslocalização de carga e o facto do anião bis(trifluorometanosulfonil)imidato ser um anião organofluorado, para além das dimensões deste anião serem semelhantes às do catião (ver discussão para sal de bis(trifluorometanosulfonil)imidato **12a** método C de síntese). No último passo de síntese foi obtido o sal de guanidínio bromado **17a** a partir do sal **16a**, em ambos os casos o anião é o bis(trifluorometanosulfonil)imidato, diferindo estes dois sais nos catiões usados, ou seja, o sal **16a** possui duas posições olefinicas que são transformadas em ligações carbono-bromo em **17a**. Nesta conversão observa-se um incremento muito significativo na densidade (1,284 vs 1,68 g/cm³ para **16a** e **17a** respectivamente) dada a elevada massa dos átomos de bromo, observando-se igualmente um incremento na temperatura de transição para o estado líquido (-56,4 vs 74,3°C para **16a** e **17a** respectivamente) dado o volume e polarizabilidade associadas ao elemento bromo (ver discussão para composto **11a** método B de síntese). Os compostos **15a**, **16a** e **17a** não apresentam quantidades de água detectáveis por análise elementar.

Comparando as propriedades dos compostos usados/obtidos no método C de síntese com as propriedades dos compostos obtidos com o método E, mais precisamente, comparando os cloretos **6a** e **15a**, pode-se observar que as suas densidades são praticamente idênticas (1,06 vs 1,02 g/cm³ para **6a**

e **15a** respectivamente) já o ponto de fusão varia significativamente (79,2 vs 146,5°C para **6a** e **15a** respectivamente) sendo mais elevado para o sal contendo o catião mais ramificado **15a**. Por sua vez, na comparação entre os sais de bis(trifluorometanosulfonil)imidato **12a** e **16a** pode-se observar um decréscimo ligeiro na densidade deste último (1,284 vs 1,342 g/cm³ para **16a** e **12a** respectivamente) sendo praticamente idênticos os valores de transição vítrea. Relativamente à densidade, as ligeiras diminuições verificadas em **15a** e **16a** relativamente a **6a** e **12a** respectivamente poderão dever-se à perda de simetria no catião. Finalmente comparando os sais bromados **13a** e **17a** pode-se observar uma diminuição significativa da densidade em **17a** (1,68 vs 2,15 g/cm³ em **17a** e **13a** respectivamente) já que neste sal foram introduzidos apenas quatro átomos de bromo enquanto que em **13a** foram introduzidos oito átomos de bromo. A elevada massa do elemento bromo provoca o aumento da densidade e sendo quatro átomos de bromo em vez de oito o aumento não é tão significativo. Relativamente ao ponto de fusão os 74,3°C relativos ao sal tetrabromado **17a** são significativamente mais baixos que a temperatura de transição para o estado líquido do sal octabromado **13a** (>143,2°C), esta diminuição drástica compreende-se dado o volume e polarizabilidade do átomo de bromo contribuírem para um acréscimo do ponto de fusão (ver discussão para composto **11a**). Tendo em conta que o ponto de fusão do sal tetrabromado **17a** é 74,3°C este pode-se considerar um líquido iónico (ponto de fusão inferior a 100°C), um líquido iónico de elevada densidade.

Seguindo o **método F** de síntese foi obtido o sal de guanidínio **18a** (**Tabela II.45**).

Tabela II.46: Transição vítrea do cloreto de guanidínio 18a obtido pelo método F de síntese.

Composto	 <p style="text-align: center;">18a</p>
Ponto de fusão (transição vítrea) (°C)	(-53,3)
Densidade (g/cm ³)	-

O sal de guanidínio **18a** representado na **Tabela II.46** apesar ter como anião o cloreto apresenta-se líquido à temperatura ambiente (transição vítrea é -53,3°C), um resultado prometedor, tendo em vista a obtenção de um sal bromado de elevada densidade líquido à temperatura ambiente, já que os precursores cloretos **6a** e **15a** eram sólidos à temperatura ambiente. O sal **18a** possui grupos substituintes de tamanhos significativamente diferentes, esta característica havia sido associada à obtenção de sais com baixo ponto de fusão no estudo **II-3.4. d)** de previsão de pontos de fusão de sais de guanidínio. Assim, estimava-se que o ponto de fusão do sal bromado contendo como anião o bis(trifluorometanosulfonil)imidato derivado de **18a** pudesse ter um ponto de fusão mais baixo do que

17a inclusivamente ser líquido à temperatura ambiente. A síntese deste hipotético composto não pôde ser concretizada já que o rendimento de síntese de **18a** é muito baixo (14%).

Como não foi possível a síntese de líquidos iônicos à temperatura ambiente de elevada densidade, no seguimento deste trabalho foram seguidas duas vias, a primeira consiste no uso destes sais de guanidínio como aditivos para aumentar a densidade de solventes de referência como é o caso do feniletanol (densidade = 1,0133 g/cm³ a 14°C), neste caso a introdução do sal **17a** como aditivo (12,5% mol. como aditivo) permitiu incremento na densidade para 1,2881 g/cm³ a 13°C. Usando o sal **13a** como aditivo (1,25% mol. como aditivo) permitiu incrementar a densidade do feniletanol para 1,1028 g/cm³ a 15°C. Outro solvente de referência é o BMIMBF₄ (tetrafluoroborato de 1-butil-3-metilimidazólio) que tendo como densidade 1,2047 g/cm³ a 13°C foi adicionado como aditivo **17a** (12,5% mol. como aditivo) sendo obtida densidade de 1,3369 g/cm³ a 12°C. Todas as soluções preparadas são homogêneas. O uso dos sais **13a** e **17a** como aditivos poderá ajudar a separação de um sistema de duas ou mais fases, em que os solventes envolvidos tenham densidades similares. A segunda via seguida, já fora do âmbito do trabalho desenvolvido neste documento, consistiu em colocar os sais densos preparados nesta secção na presença de CO₂ super-crítico de forma a permitir a diminuição do ponto de fusão desses sais já que em estudo conduzido por Scurto et. al.¹¹¹ verificou-se o decréscimo do ponto de fusão de alguns sais de imidazólio, piridínio, amónio, sulfónio e fosfónio. Nesse sentido usou-se o mesmo princípio para provocar a diminuição do ponto de fusão de sais de guanidínio. Este trabalho não foi desenvolvido por mim, foi antes uma colaboração com o grupo de processos com fluidos super-críticos liderado pelo Professor Manuel Nunes da Ponte. Nesse trabalho foi testado o sal **17a** (*p.f.* = 82-85°C medido no sistema de CO₂ super-crítico na ausência de CO₂) que na presença de 113 bar de CO₂ super-crítico permitiu a diminuição do ponto de fusão para 44,6°C, desta forma é possível ter um sistema binário CO₂ super-crítico/líquido iónico denso **17a** a partir de 44,6°C.

II-3.6 - Estudo F: Síntese de sais de Guanidínio contendo Tiocianato, *para*-Toluenosulfonato ou Dicianamida como Aniões e Estudo alargado da Influência de Catião e Anião no Ponto de Fusão

Na secção II-3.4 d) foram construídos dois modelos que explicavam e previam o ponto de fusão de sais contendo catiões baseados na unidade de guanidínio e contendo como anião o tetrafenilborato, cloreto, brometo ou iodeto. Com o auxílio desses modelos foi possível por evidência alguns padrões estruturais relacionados principalmente com a estrutura do catião, nomeadamente com as cadeias substituintes do núcleo guanidínio. Adicionalmente este modelo indicou novos sais de guanidínio (cloretos e tetrafenilboratos) com baixo ponto de fusão e que foram sintetizados laboratorialmente. Apesar da utilidade revelada por estes dois modelos, baseados no algoritmo de rede neuronal de *Counter-Propagation*, o seu uso encontrava-se limitado a cloretos, brometos, iodetos e tetrafenilboratos, aniões presentes em quantidade mais abundante na literatura quando associados a catião baseado em núcleo de guanidínio. Estes aniões cuja carga negativa é pouco ou nada deslocalizável, conduzem tradicionalmente a sais com ponto de fusão relativamente elevado, salvo escasso número de excepções identificadas e abordadas na secção II-3.4 d). O estudo contido nesta secção foi realizado como extensão de estudo anterior II-3.4 d) já que consiste na síntese e estudo de novos sais de guanidínio contendo como aniões o tiocianato, o *para*-toluenosulfonato, dicianamida (**Figura II.35**) aniões com grande capacidade estrutural para deslocalizar a carga negativa logo com potencial para a obtenção de novos líquidos iônicos à temperatura ambiente ou temperatura próxima.¹¹²

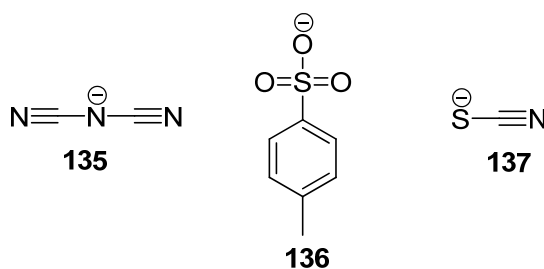


Figura II.35: Aniões usados na síntese de novos sais de guanidínios – 135 dicianamida, 136 *para*-toluenosulfonato e 137 tiocianato.

Adicionalmente foram considerados sais contendo os aniões perfluorooctanoato, hexafluorofosfato, bis(trifluorometanosulfonyl)imidato, cloreto e tetrafenilborato, sintetizados para os estudos II-3.4 d) e II-3.5 e) de forma a estudar o efeito integrado de catião e anião no ponto de fusão, principalmente o efeito do anião.

Para a obtenção dos sais baseados nos aniões tiocianato, dicianamida e *para*-toluenosulfonato, sintetizados especificamente para este estudo, partiu-se dos seguintes sais de guanidínio contendo cloreto como anião e que foram sintetizados para o estudo II-3.4 d) (Figura II.36):

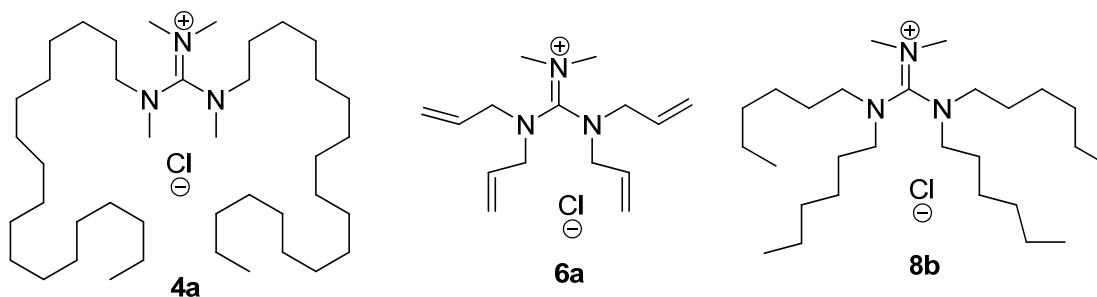


Figura II.36: Cloretos de guanidínio sintetizados para o estudo d) e usados na síntese de novos sais de guanidínio contendo tiocianato, *para*-toluenosulfonato e dicianamida como aniões.

Estes três compostos de partida, tal como foi discutido na secção II-3.4 d), possuem elevada pureza ainda que contenham quantidade considerável de água associada às suas estruturas (Tabelas II.29 e II.30). O cloreto **4a** foi submetido às seguintes trocas de anião:

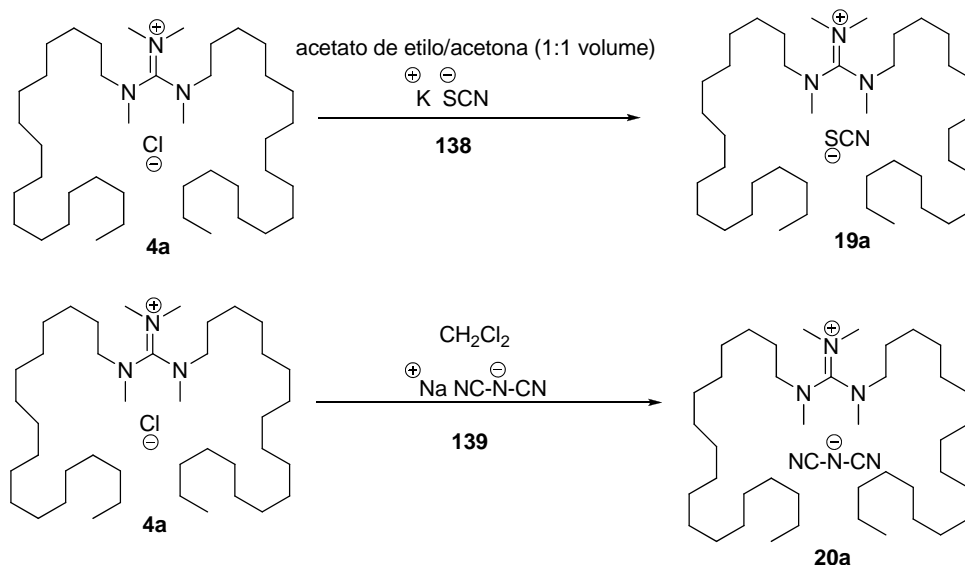
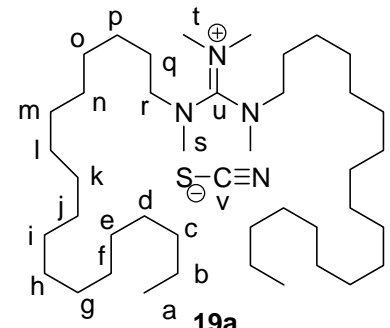


Figura II.37: Trocas de anião para a obtenção dos sais baseados nos aniões tiocianato **19a** e dicianamida **20a**.

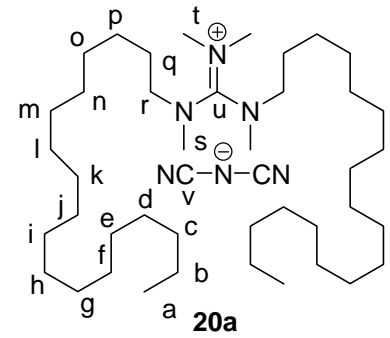
Os dois sais sintetizados, **19a** e **20a** foram caracterizados espectroscopicamente e por análise elementar (Tabela II.47) de forma a avaliar se a troca de anião foi completa (catião e anião pretendido na proporção 1:1) e se cada composto sintetizado está puro.

Tabela II.47-1: Dados reaccionais, espectrais e análise elementar do sal baseado no anião tiocianato 19a e dicianamida 20a.

	Tempo de reacção η (%)	^1H -RMN δ (ppm) J (Hz)	^{13}C -RMN δ (ppm)	IV(filme ou KBr) $\nu_{\text{máx}}$ (cm $^{-1}$)	Análise elementar
	48 horas 83,6%	0,83-0,87 (6H, t, $J = 6,6$) Ha 1,23 (64H, complexo) Hb-q 2,99-3,11 (12H, complexo) Hs,t 3,23 (4H, m) Hr	14,02 Ca 22,59 Cb 26,74 Cp 27,62 Cq 29,20 29,27, 29,62, 31,84 Cc-o 38,12, 40,36, 40,78 Cs,t 52,87 Cr 163,68 Cu	IV(KBr) 2920 (CH alifática) 2850 (CH alifática) 2060 (S=C=N) 1585 (C-N) 1562 (C-N)	composição calculada para $\text{C}_{42}\text{H}_{86}\text{N}_4\text{S}$ (%): N 8,25, C 74,27, H 12,76; composição determinada (%): N 8,39, C 74,30, H 12,92.

η – rendimento (%), δ - desvio químico, J – constante de acoplamento, $\nu_{\text{máx}}$ – energia representada em número de onda.

Tabela II.47-2:Dados reaccionais, espectrais e análise elementar do sal baseado no anião tiocianato 19a e dicianamida 20a.

	Tempo de reacção	¹ H-RMN δ (ppm) J (Hz)	¹³ C-RMN δ (ppm)	IV(filme ou KBr) ν _{máx} (cm ⁻¹)	Análise elementar
	24 horas	0,86-0,90 (6H, t, J = 6,4) Ha	14,06 Ca	IV(KBr)	composição calculada para C₄₃H₈₆N₆ (%): N 12,23, C 75,16, H 12,61; composição determinada (%): N 12,23, C 74,99, H 12,83.
	67,1%	1,26 (64H, complexo) Hb-q 2,97-3,09 (12H, complexo) Hs,t 3,21 (4H, m) Hr	22,63 Cb 26,75 Cp 27,64 Cq 29,21, 29,31, 29,53, 29,66 31,87 Cc-o 38,11, 38,38, 40,36, 40,73 Cs,t 52,93 Cr 119,93 Cv 163,65 Cu	2920 (CH alifática) 2850 (CH alifática) 2225 (NC-N-CN) 2187 (NC-N-CN) 2133 (NC-N-CN) 1570 (C-N)	

η – rendimento (%), **δ** - desvio químico, **J** – constante de acoplamento, **ν_{máx}** – energia representada em número de onda.

Percorrendo a **Tabela II.47** é possível observar pela análise aos dados dos espectros de ¹H-RMN dos dois compostos sintetizados que está presente o catião pretendido, esse facto é reforçado pelos dados dos espectros de ¹³C-RMN em especial o sinal **Cu** a *ca* de 163 ppm; relativamente à presença do anião pretendido no caso do composto **19a** esta não pode ser confirmada por ¹³C-RMN já que não aparece nenhum sinal no espectro relativamente ao carbono **Cv**, este facto pode-se explicar pelo facto de haver apenas um carbono **Cv** por unidade de composto adicionado ao facto de não haver nenhum hidrogénio acoplado a esse carbono. A presença do anião tiocianato em **19a** só pôde ser confirmada pelos dados da análise elementar, já que há correspondência entre a composição determinada e a composição calculada em que aparece incluído o anião tiocianato. A indicação da presença deste anião é reforçada pelos dados do espectro de Infravermelho já que aparece banda muito intensa a 2060 cm⁻¹ que não aparecia no espectro de Infravermelho do respectivo cloreto **4a** (**Tabela II.29**). Analisando as outras bandas do espectro de Infravermelho de **19a** confirmam a presença do catião pretendido. Relativamente ao composto **20a**, os dados do respectivo espectro de ¹³C-RMN indicam a presença do anião dicianamida já que aparece um sinal a 119,93 ppm correspondente ao carbono **Cv** do anião, esse dado é confirmado por análise elementar já que os dados da composição determinada correspondem aos dados da composição calculada em que aparece incluído o anião dicianamida (na proporção 1:1 relativamente ao catião). Os dados do espectro de Infravermelho do composto **20a** também confirmam a presença de anião dicianamida já que aparecem

bandas intensas na gama 2225-2133 cm^{-1} que não aparecem no espectro do cloreto respectivo **4a** (Tabela II.29). O espectro de Infravermelho de **20a** confirma a presença das funcionalidades/ligações do catião pretendido.

Usando estratégia semelhante à usada na síntese dos compostos **19a** e **20a** foram sintetizados novos sais baseados igualmente no anião dicianamida **21a** e tiocianato **22a**, para estas duas sínteses partiu-se do cloreto **6a** participando em duas reacções de troca de anião (Figura II.38).

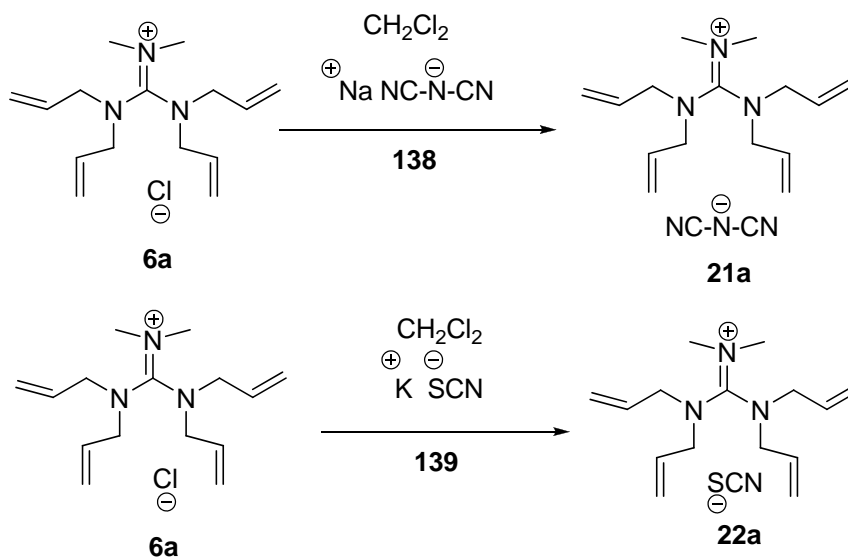


Figura II.38: Reacções de troca de anião para a síntese dos sais de dicianamida **21a** e tiocianato **22a**.

De acordo com **Figura II.38** foram sintetizados os sais de guanidínio baseados no anião dicianamida **21a** e tiocianato **22a** que foram caracterizados espectroscopicamente e através de análise elementar (Tabela II.48) de forma a inferir se a troca de anião foi completa, ou seja saber se o catião e o anião estão na proporção 1:1, e se os compostos obtidos possuem elevado grau de pureza.

Tabela II.48: Dados reaccionais, espectroscópicos e análise elementar do sal baseado no anião dicianamida 21a e tiocianato 22a.

	Tempo de reacção η (%)	^1H -RMN δ (ppm) J (Hz)	^{13}C -RMN δ (ppm)	IV(filme ou KBr) $\nu_{\text{máx}}$ (cm^{-1})	Análise elementar
<p>21a</p>	72 horas 73,13%	3,00 (6H, s) Ha 3,77-3,85 (8H, m) Hd 5,32-5,36 (8H, m) Hb 5,74 (4H, sl) Hc	40,48 Ca 52,44, 52,87 Cd 119,82 Cf 121,89, 122,08 Cb 130,57 Cc 162,97 Ce	IV (KBr) 3086 (CH olefínica) 3000 (CH alifática) 2926 (CH alifática) 2293, 2229, 2195 e 2137 (NC-N-CN) 1644 (C-N) 1585 (C-N) 1538 (C-N) 996 (CH olefínica) 931 (CH olefínica) 880 (CH olefínica)	composição calculada para $\text{C}_{17}\text{H}_{26}\text{N}_6$ (%): N 26,73, C 64,94, H 8,33; composição determinada (%): N 26,75, C 64,64, H 8,86.
<p>22a</p>	72 horas 58,82%	3,07 (6H, s) Ha 3,79-3,96 (8H, m) Hd 5,31-5,34 (8H, m) Hb 5,79 (4H, sl) Hc	40,65 Ca 52,49, 52,83 Cd 121,68 Cb 130,98 Cc 163,01 Ce	IV(KBr) 3086 (CH olefínica) 3000 (CH alifática) 2914 (CH alifática) 2059 (SCN) 1664 (C-N) 1585 (C-N) 1539 (C-N) 993 (CH olefínica) 933 (CH olefínica) 882 (CH olefínica) 868 (CH olefínica)	composição calculada para $\text{C}_{16}\text{H}_{26}\text{N}_4\text{S}_0,1\text{H}_2\text{O}$ (%): N 18,17, C 62,34, H 8,57 composição determinada (%): N 18,60 C 62,29, H 8,79

η – rendimento (%), δ - desvio químico, J – constante de acoplamento, $\nu_{\text{máx}}$ – energia representada em número de onda.

De acordo com os dados da **Tabela II.48** os compostos sintetizados, **21a** e **22a** foram obtidos com elevado grau de pureza. Concretamente, através da análise aos espectros ^1H -RMN destes compostos pode-se concluir que o catião é o pretendido em ambos os casos, esse facto é reforçado pelos dados dos espectros de ^{13}C -RMN havendo correspondência entre cada um dos sinais dos espectros e os diferentes tipos de carbonos, nestes espectros há a destacar os sinais a *ca* de 163 ppm correspondentes ao carbono quaternário Ce característico de sais de guanidínio. No caso do composto **21a** o espectro de ^{13}C -RMN indica a presença de dicianamida aparecendo sinal a 119,82 ppm correspondendo a carbono Cf do anião. A presença deste anião é confirmada por análise elementar já que há correspondência entre a composição determinada experimentalmente e a composição calculada incluída a presença do anião dicianamida, esse dado confirma a pureza do composto. Ainda

relativamente ao composto **21a**, pela análise ao espectro de Infravermelho pode-se confirmar a presença de anião pela presença de bandas intensas na gama $2137\text{--}2293\text{ cm}^{-1}$ que não apareciam no cloreto precursor **6a** (Tabela II.30), a análise às restantes bandas do espectro de IV confirmam também a presença das funcionalidades/tipos de ligação presentes no catião pretendido. Relativamente ao composto **22a** os dados de ^{13}C -RMN não confirmam a presença de anião tiocianato pelo mesmo motivo apresentado para o composto **19a**, o anião é sim confirmado por análise elementar já que há correspondência entre composição determinada experimentalmente e composição calculada tendo em conta a presença de anião tiocianato na proporção 1:1 em relação ao catião; estes dados confirmam a elevada pureza do composto **22a**. Relativamente ao espectro de Infravermelho este reforça a indicação em favor da presença do anião pretendido aparecendo banda característica a 2059 cm^{-1} que não aparece no espectro do cloreto precursor **6a** (Tabela II.30). As restantes bandas do espectro de Infravermelho confirmam a presença das funcionalidades/tipos de ligação do catião pretendido.

O último composto sintetizado desta série de sais baseia-se no anião *para*-toluenosulfonato que foi obtido segundo reacção de troca de anião partindo de cloreto **8b** a reagir na presença de *para*-toluenosulfonato de sódio (Figura II.39).

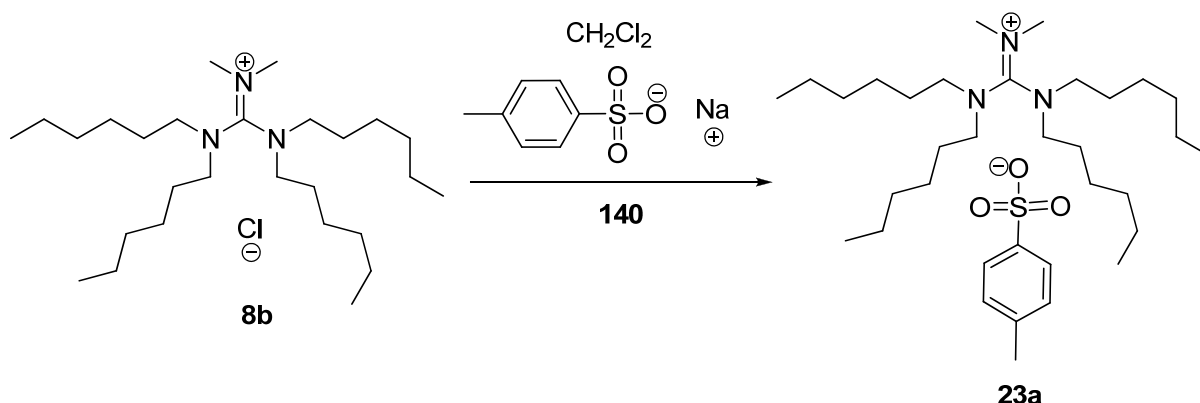
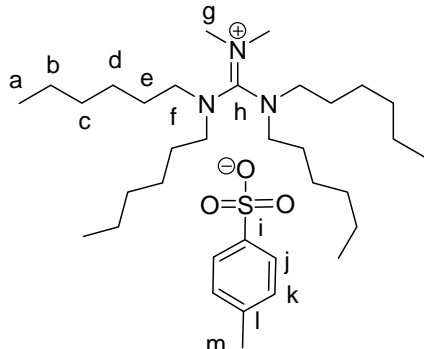


Figura II.39: Reacção de troca de anião para a síntese do sal contendo anião *para*-toluenosulfonato **23a**.

Segundo a Figura II.39 foi sintetizado o sal baseado no anião *para*-toluenosulfonato **23a** e que foi caracterizado (Tabela II.49) de forma a demonstrar que a reacção de troca de anião foi completa e a proporção entre catião e anião é 1:1 e que o composto sintetizado tem elevado grau de pureza.

Tabela II.49: Dados reaccionais, espectrais e análises elementares relativamente ao sal contendo anião *para*-toluenosulfonato 23a.

	Tempo de reação η (%)	^1H -RMN δ (ppm) J (Hz)	^{13}C -RMN δ (ppm)	IV(filme ou KBr) $\nu_{\text{máx}}(\text{cm}^{-1})$	Análise elementar
 <p>23a</p>	24 horas 98%	0,78-0,85 (12H, m) Ha 1,22-1,66 (32H, complexo) Hb-e 2,29 (3H, s) Hm 3,03 (6H, s) Hg 3,08-3,17 (8H, m) Hf 7,07-7,09 (2H, d, J = 7,7) Hk 7,73-7,74 (2H, d, J = 7,7) Hj	13,84 Ca 21,15 Cm 22,39 Cb 26,34, 26,47 Cd 27,36, 27,69 Ce 31,22, 31,39 Cc 40,71 Cg 49,27, 49,80 Cf 125,98 Cj 128,30 Ck 138,83 Ci 143,72 Cl 163,34 Ch	IV(filme) 2927 (CH alifática) 2858 (CH alifática) 1643 (C-N) 1581 (C-N) 1539 (C-N) 1219 (RSO ₃ ⁻) 818 (CH aromática) 729 (CH aromática)	composição calculada para C₃₄H₆₅N₃O₃S (%) : N 7,05, C 68,52, H 10,99 composição determinada (%) : N 6,87 C 68,26 H 11,46

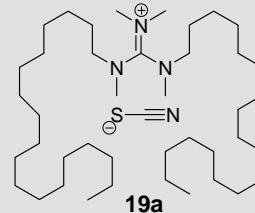
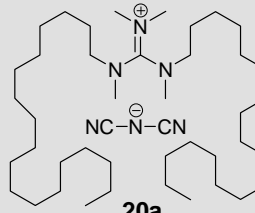
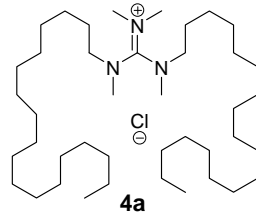
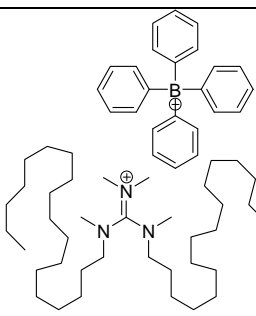
η – rendimento (%), δ - desvio químico, J – constante de acoplamento, $\nu_{m\acute{a}x}$ – energia representada em número de onda.

Tendo em conta os resultados apresentados na **Tabela II.49** pode-se concluir que o composto **23a** foi obtido com elevado grau de pureza já que analisando o espectro de ^1H -RMN pode-se observar a correspondência entre os desvios químicos apresentados (tal como a multiplicidade) e os tipos de prótons presentes na molécula alvo. Adicionalmente pode-se observar que catião e anião se encontram na proporção 1:1, por exemplo através da comparação das áreas de **Hg** e **Hk** do catião e anião respectivamente. Os dados de pureza do composto foram confirmadas por análise ao espectro de ^{13}C -RMN, neste caso há correspondência entre os desvios químicos obtidos e os tipos de carbono presentes na estrutura alvo. A destacar neste espectro o desvio químico a 163,34 ppm característico de carbono quaternário **Ch** de núcleo guanidínio; pode igualmente destacar-se os desvios químicos dos carbonos aromáticos do anião na gama 125,98-143,72 ppm correspondentes aos carbonos **Ci-l** sinal da presença do anião pretendido no composto sintetizado. Relativamente ao espectro de Infravermelho é possível identificar todas as funcionalidades/tipos de ligações presentes no composto alvo **23a**, concretamente foram identificadas claramente bandas a 2858 e 2927 cm^{-1} características de ligações CH alifáticas, por outro lado observou-se bandas a 1643, 1581 e 1539 cm^{-1} típicas de ligações CN de núcleo guanidínio, por sua vez o anião foi identificado através de banda característica a 1219 cm^{-1} típica da funcionalidade sulfonato do anião, este foi ainda identificado por bandas a 818 e 729 cm^{-1}

típicas de ligações CH aromáticas. Para poder concluir que o composto estava definitivamente puro recorreu-se a análise elementar e esta permitiu concluir que o composto estava efectivamente puro já que a composição determinada experimentalmente para a amostra analisada corresponde à composição calculada para a estrutura **23a** alvo.

Posteriormente, os novos compostos sintetizados nesta secção foram comparados relativamente ao ponto de fusão experimental juntamente com outros compostos sintetizados para os estudos das secções II-3.4. d) e II-3.5. e). Desta forma foi comparada a série de sais baseada no catião *N,N,N',N''*-tetrametil-*N',N''*-dioctadecilguanidínio:

Tabela II.50: Comparação entre diversos sais baseados no catião de *N,N,N',N''*-tetrametil-*N',N''*-dioctadecilguanidínio relativamente ao ponto de fusão.

Composto	 19a	 20a	 4a	 7a
Ponto de fusão (°C)	83	78	78	88

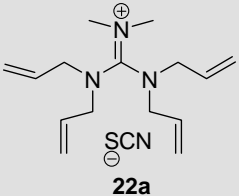
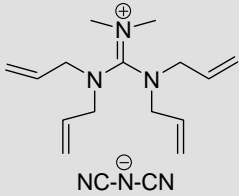
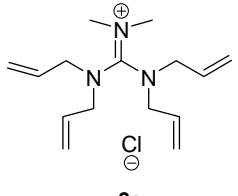
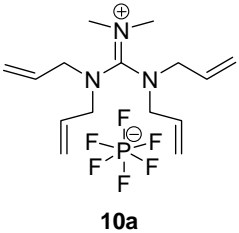
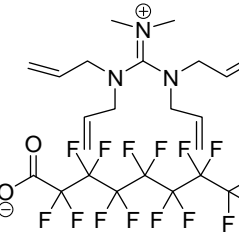
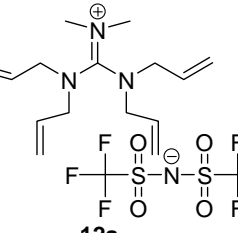
Com fundo cinzento aparecem assinalados os sais sintetizados especificamente para este estudo.

Tendo em conta os resultados apresentados na **Tabela II.50** para quatro sais baseados no mesmo catião de *N,N,N',N''*-tetrametil-*N',N''*-dioctadecilguanidínio mas variando o anião é possível observar que, apesar dos aniões em causa serem muito diversificados tanto ao nível estrutural como ao nível de volume, o ponto de fusão varia muito pouco, desde 78°C, para o cloreto **4a** e sal de dicianamida **20a**, passando pelos 83°C do tiocianato **19a**, até aos 88°C do sal de tetrafenilborato **7a**. Estes resultados indicam que o ponto de fusão deste tipo de sais é pouco influenciado pelo anião. Ainda assim pode-se explicar o facto do sal de tetrafenilborato **7a** ter um ponto de fusão mais elevado que os restantes sais pelo volume do anião tetrafenilborato ser relativamente semelhante ao volume do catião, reduzindo assim a entropia de fusão, este anião também não permite a deslocalização da carga negativa por ressonância factor que faz aumentar o ponto de fusão. Por sua vez o tiocianato e a dicianamida permitem deslocalização de carga, enquanto que o cloreto, ainda que não permita essa deslocalização de carga, é o anião mais pequeno entre os aniões estudados, conduzindo a um ponto de fusão mais baixo, neste caso o factor entrópico também está em evidência. Neste caso foi possível a

obtenção de novos líquidos iônicos (ponto de fusão inferior a 100°C) usando o tiocianato e a dicianamida como aniões muito embora estes compostos não sejam líquidos à temperatura ambiente.

No seguimento deste estudo foi igualmente comparado o comportamento de sais baseados no catião *N,N*-dimetil-*N',N',N'',N''*-tetralilguanidínio, nesta comparação está-se a estudar a variação estrutural no anião, sendo estudados seis aniões diferentes (**Tabela II.51**).

Tabela II.51: Comparação entre diversos sais baseados no catião *N,N*-dimetil-*N',N',N'',N''*-tetraalilguanidínio relativamente ao ponto de fusão.

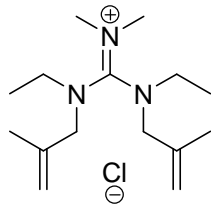
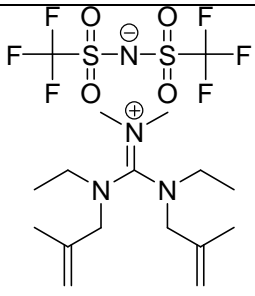
Composto	 22a	 21a	 6a
Ponto de fusão (°C)	71,5	50	79
Composto	 10a	 14a	 12a
Ponto de fusão (transição vítrea) (°C)	56	(-49,4)	(-58)

Com fundo cinzento aparecem assinalados os sais sintetizados especificamente para este estudo.

Tendo em conta os resultados apresentados na **Tabela II.51** é possível observar uma variação mais pronunciada do ponto de fusão com o anião utilizado em relação aos casos apresentados na **Tabela II.50**. Usando na comparação os sais baseados nos aniões dicianamida **21a**, tiocianato **22a** e cloreto **6a**, aniões que também foram usados nos exemplos da **Tabela II.50**, pode-se ver que agora a variação entre pontos de fusão é maior, desde os 50°C do sal baseado no anião dicianamida, passando pelos 71,5°C do sal baseado no anião tiocianato, até aos 79°C do cloreto. Nesta comparação foram ainda considerados os sais baseados nos aniões hexafluorofosfato **10a**, bis(trifluorometanosulfonyl)imidato **12a**, e perfluorooctanoato **14a**. Estes dois últimos sais são líquidos iônicos à temperatura ambiente apresentando valores de transição vítrea baixos (-49,4°C para

o sal baseado no anião perfluorooctanoato **14a** e -58°C para o sal baseado no anião bis(trifluorometanosulfonil)imidato **12a**). Para estes dois aniões a quantidade de água presente é 0,98 % e 3,1% em massa respectivamente para o caso dos aniões bis(trifluorometanosulfonil)imidato e perfluorooctanoato, sendo para este último caso um valor significativo, assim há que olhar com as devidas cautelas para a temperatura de transição para o estado líquido associada a este composto. Estes dois aniões para além de terem a carga negativa deslocalizada são aniões organo-fluorados sendo tipos de estruturas que apresentam interações de baixa intensidade com outras estruturas orgânicas (neste caso o catião guanidínio), isto já que nas estruturas organo-fluoradas, devido a elevada electronegatividade do flúor, os carbonos associados apresentam carga parcial positiva algo que não acontece com as habituais estruturas orgânicas, adicionalmente estruturas organo-fluoradas são muito rígidas constituindo diversos confórmeros aumentando a entropia do sistema,¹¹³ estes factos somados podem permitir a redução do ponto de fusão. Tendo em conta os resultados avançados o perfluorooctanoato é um anião promissor tendo em conta a obtenção de novos líquidos iônicos à temperatura ambiente. Foram assim obtidos seis novos líquidos iônicos, incluídos os sais de tiocianato **22a** e dicianamida **21a**, preparados especificamente para este estudo, neste conjunto de seis compostos estão ainda incluídos dois novos líquidos iônicos à temperatura ambiente.

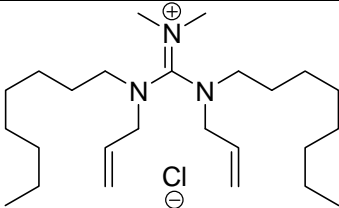
Tabela II.52: Comparação entre dois sais baseados no catião *N,N*-dimetil-*N',N''*-dietil-*N',N''*-bis(2-metilalil)guanidínio ao nível da transição para o estado líquido.

Composto	 15a	 16a
Ponto de fusão (transição vítrea) (°C)	146,5	(-56,4)

Seguindo mesma Filosofia, na **Tabela II.52**, foram comparados os pontos de fusão de dois novos sais baseados no catião *N,N*-dimetil-*N',N''*-dietil-*N',N''*-bis(2-metilalil)guanidínio variando o anião utilizado em cada caso. Assim, foi estudado o sal **15a**, baseado no anião cloreto, e o sal **16a** baseado no anião bis(trifluorometanosulfonil)imidato, claramente este último apresenta transição mais baixa para o estado líquido, o anião utilizado é organo-fluorado com carga negativa deslocalizada. O teor em água associado ao composto **16a** é 0,80% em massa. Comparando o sal **15a** com o composto **6a** (**Tabela II.51**), é possível observar que para aquele o ponto de fusão é significativamente superior (146,5°C vs 79°C para **6a**), mesmo tratando-se de sais com o mesmo anião (cloreto), e possuindo ambos catiões de dimensões similares, apresentando estes grupos olefinicos, o que os distingue é que

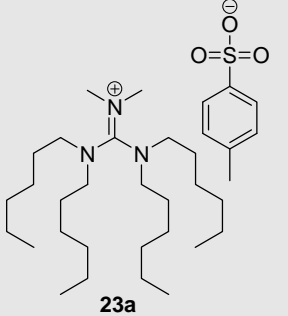
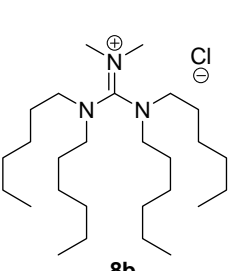
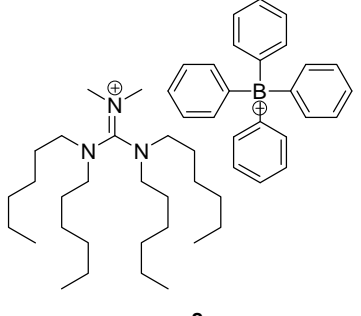
no sal **15a** existem apenas dois grupos olefinicos (enquanto que em **6a** existem quatro grupos olefinicos) e em **15a** a estrutura é ramificada, algo que em **6a** não acontece. Neste caso sabe-se que a presença de ramificações contribui para o aumento do ponto de fusão pela perda de graus de liberdade conformacionais. Nesta comparação entre cloretos, verifica-se que a influência do catião é muito significativa na definição do ponto de fusão. Esta observação é contrastada pela comparação entre os sais de bis(trifluorometanosulfonil)imidato **16a** e **12a** (Tabela II.51) sais cujo valor de transição vítrea é praticamente idêntico (-56,4 vs -58°C para **12a**), neste caso a variação no catião praticamente não conduz a variação na transição para o estado líquido.

Tabela II.53: Cloreto de *N,N*-dimetil-*N',N''*-dialil-*N',N''*-dioctilguanidínio e respectiva transição vítrea.

Composto	 <p>18a</p>
(Transição vítrea) (°C)	(-53,3)

Na mesma sequência foi preparado o cloreto de *N,N* - dimetil - *N',N''*-dialil-*N',N''*-dioctilguanidínio (Tabela II.53), composto baseado em catião que, à semelhança dos casos das Tabelas II.51 e II.52, possui grupos olefinicos, este catião diverge dos casos anteriores pela assimetria entre os tamanhos das cadeias, havia ficado evidente no estudo II-3.4. d) que esta característica conduzia a compostos com baixo ponto de fusão e foi o que aconteceu com o composto **18a** possuindo uma transição vítrea de -53,3°C, valor de transição para o estado líquido significativamente mais baixo que os cloretos **6a** e **15a** (Tabelas II.51 e II.52 respectivamente) compostos cujos grupos substituintes à volta do núcleo guanidínio são mais semelhantes em termos de dimensão.

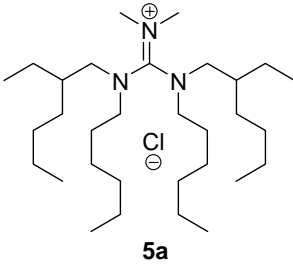
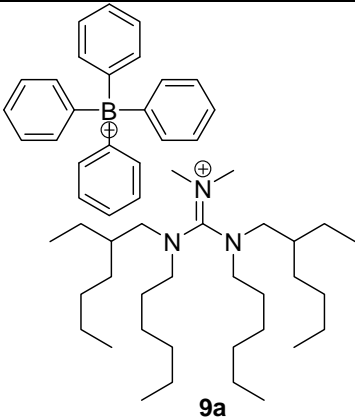
Tabela II.54: Comparação entre diversos sais baseados no catião *N,N*-dimetil-*N'*,*N''*,*N'''*-tetrahexilguanidíneo ao nível da transição para o estado líquido.

Composto	 23a	 8b	 8a
Ponto de fusão (transição vítrea) (°C)	(-66,5)	(-63)	58

Com fundo cinzento aparece assinalado o sal sintetizado especificamente para este estudo.

Relativamente aos compostos representados na **Tabela II.54**, baseiam-se todos no catião *N,N*-dimetil-*N'*,*N''*,*N'''*-tetrahexilguanidíneo variando de caso para caso o anião utilizado, observando-se a presença de dois líquidos iônicos à temperatura ambiente, o sal baseado no anião *para*-toluenosulfonato **23a**, e o cloreto **8b**, possuindo transições vítreas a -66,5°C e -63°C respectivamente. Por sua vez o composto **8a** baseado no anião tetrafenilborato apresenta como ponto de fusão 58°C, já se havia observado também no composto **7a** (**Tabela II.50**) que este anião conduzia a líquidos iônicos com pontos de fusão mais elevados, sendo este um anião sem capacidade para deslocalizar a carga negativa por ressonância ao mesmo tempo que tem dimensões próximas aos catiões em **7a** e **8a**, características que conduzem a pontos de fusão mais elevados. O anião *para*-toluenosulfonato permite uma grande deslocalização da carga negativa por ressonância conduzindo à obtenção de um novo líquido iónico à temperatura ambiente (**23a**), o anião cloreto possui dimensão substancialmente mais reduzida que o catião conduzindo a entropia de fusão suficientemente baixa para a obtenção de um líquido iónico à temperatura ambiente. Comparando os sais baseados nos aniões cloreto e tetrafenilborato da **Tabela II.50** (**4a** e **7a**) e **Tabela II.54** (**8a** e **8b**), pode-se observar que neste caso a transição para o estado líquido varia significativamente com o anião (-63°C para o cloreto e 58°C para o sal de tetrafenilborato) enquanto que para o catião representado na **Tabela II.50** a variação entre cloreto e sal de tetrafenilborato é praticamente negligenciável (78 vs 88°C respectivamente), ou seja, a estrutura do catião influencia significativamente o impacto que a variação entre aniões vai ter na transição para o estado líquido. Neste estudo foi sintetizado um novo líquido iónico à temperatura ambiente, o sal baseado no anião *para*-toluenosulfonato **23a**, o também líquido iónico à temperatura ambiente **8b** já havia sido sintetizado por Mateus *et al*⁴² anteriormente.

Tabela II.55: Comparação entre diversos sais baseados no catião *N,N*-dimetil-*N',N''*-dihexil-*N',N''*-bis(2-etilhexil)guanidínio ao nível da transição para o estado líquido.

Composto	 <p style="text-align: center;">5a</p>	 <p style="text-align: center;">9a</p>
Ponto de fusão (transição vítrea) (°C)	(-86)	125,6

Finalmente, analisando a **Tabela II.55** pode-se observar dois sais baseados no catião *N,N*-dimetil-*N',N''*-dihexil-*N',N''*-bis(2-etilhexil)guanidínio que foram preparados especificamente para o estudo **II-3.4. d)**. Neste caso está-se a estudar a variação entre anião cloreto e tetrafenilborato. Assim, à semelhança do cloreto **8b** e sal de tetrafenilborato **8a** da **Tabela II.54**, há grande variação na transição para o estado líquido entre o cloreto **5a** e o sal de tetrafenilborato **9a** (-86°C vs 125,6°C respectivamente) sendo mais marcada a variação no caso deste catião (**Tabela II.55**) diferindo este do catião representado na **Tabela II.54** pela presença de dois grupos etilo como ramificações naquele catião. Desta forma, à semelhança do caso da **Tabela II.54** e contrariamente aos casos da **Tabela II.50**, para este catião (**Tabela II.55**) o papel do anião na modelação do ponto de fusão é muito relevante. Geralmente a presença de ramificações aparece associada a pontos de fusão mais elevados e essa generalização é válida na comparação entre os sais de tetrafenilborato **9a** (**Tabela II.55**) e **8a** (**Tabela II.54**). Essa tendência é contrariada na comparação entre os respectivos cloretos **5a** e **8b** (transições vítreas iguais a -86°C e -63°C respectivamente), muito embora se esteja a falar de transição vítrea e não de ponto de fusão. Até agora seguiu-se uma abordagem mais centrada na influência do anião na transição para o estado líquido, estando o peso dessa influência estreitamente interligado com a estrutura do catião. Olhando um pouco mais em detalhe para a influência do catião, olhando para os cloretos das **Tabelas II.50-II.55**, pode-se observar as mesmas tendências evidenciadas em modelos baseados em redes neuronais de *Counter-Propagation* no estudo **II-3.4. d)** já que catiões com grupos substituintes à volta do núcleo guanidínio similares em tamanho apresentam as transições para o estado líquido mais elevadas (**6a** 79°C e **15a** 146,5°C) do que o caso de catiões com tamanhos de grupos substituintes mais diversificados (**18a** -53°C, **8b** -63°C e **5a** -86°C), neste caso a única exceção é o composto **4a**, que apresenta ponto de fusão igual a 78°C, neste caso dois dos grupos substituintes do catião são cadeias *n*-octadecilo, cadeias muito longas que

permitem que as interações de Van der Waals passem a ter um peso muito grande na definição do ponto de fusão.

II-4: Conclusão

De acordo com os resultados discutidos anteriormente é possível concluir que no estudo **a)** de modelação do ponto de fusão de sais de imidazólio o modelo a apresentar melhor capacidade de modelação e previsão foi um modelo RF responsável por $R^2 = 0,985$ entre valores experimentais e valores previstos de ponto de fusão para objectos do conjunto de treino. Este modelo apresenta moderada capacidade de previsão, $R^2 = 0,747$ (RMS = 28,330 °C) para conjunto independente de teste. O modelo obtido, baseado essencialmente em sólidos à temperatura ambiente, não é o mais indicado para a descoberta de novos líquidos iónicos.

Noutro estudo (**b**) foi construído modelo de classificação SVM baseado tanto em sólidos como líquidos iónicos à temperatura ambiente sendo obtida concordância = 0,984 para conjunto de treino. Este modelo apresenta moderada capacidade de previsão (concordância = 0,798 para conjunto de teste). Outros métodos como as RF e Árvores de Decisão providenciam informação complementar em relação ao modelo SVM. No caso de modelo RF medidas de probabilidade e proximidade foram obtidas tendo ficado demonstrado que ambas as medidas dão indicação correcta se a previsão obtida pode ser acertada. O modelo Árvore de Decisão permitiu extrair quais as características estruturais mais relevantes a influenciar o ponto de fusão/estado físico, neste caso características como o número de átomos e carga máxima revelaram-se as características mais relevantes para o anião. Relativamente ao catião, características como a dimensão, presença de átomos polarizáveis, ramificações e tendência para formar pontes de hidrogénio foram as seleccionadas. Novos sais de imidazólio (líquidos iónicos) foram sintetizados de acordo com as indicações dos modelos construídos. O ponto de fusão/estado físico destes sais corresponde em grande medida às previsões obtidas pelos modelos de classificação.

No estudo **c)** de densidade de sais de imidazólio foi construído modelo SVM apresentando $R^2 = 1,000$ para objectos do conjunto de treino. Este modelo apresenta elevada capacidade de previsão ($R^2 = 0,975$ e RMS = 0,059 g/cm³). Relativamente às características estruturais mais relevantes na definição da densidade, encontradas por Árvore de Decisão, foram seleccionadas a presença de átomos pesados como o bromo ou o iodo no catião, a diferença de tamanho entre catião e anião e presença de ramificações e átomos electronegativos como o flúor no anião.

No estudo **d)** relativo ao ponto de fusão de sais de guanidínio, foi obtido modelo baseado num *ensemble* de cinco redes neuronais CPG, apresentando elevada capacidade de previsão, sendo obtido $R^2 = 0,815$ (RMS = 23,950 °C) com conjunto independente de teste. Através do mapeamento de todos os objectos numa rede neuronal CPG foi possível observar alguns padrões estruturais relacionados com o ponto de fusão, concretamente observou-se que catiões de guanidínio a apresentar posições não substituídas e/ou possuindo grupos pequenos de pequenas dimensões estão associados a elevados pontos de fusão. Relativamente a estruturas de catião, grupos substituintes de tamanhos similares e/ou

apresentando elevada simetria aparecem associados a pontos de fusão intermédios. Finalmente estruturas de catião possuindo grupos substituintes de diferentes dimensões aparecem associadas a baixos pontos de fusão, algumas constituindo sais líquidos à temperatura ambiente. Com base nas previsões obtidas por *ensemble* de cinco redes CPG, foram sintetizados seis novos sais de guanidínio cujo ponto de fusão previsto aproxima-se ao ponto de fusão experimental. Foi possível sintetizar novos líquidos iónicos (líquidos à temperatura ambiente) um dos quais o sal com pf mais baixo entre todos os conhecidos para aquele anião (tetrafenilborato). Os modelos construídos estão limitados a brometos, cloretos, iodetos e a sais de tetrafenilborato, estando igualmente limitados a catiões contidos no domínio estrutural do conjunto de treino (catiões baseados em núcleo guanidínio contendo seis posições representadas por hidrogénios, cadeias alquílicas não funcionalizadas ou incluindo éteres cíclicos, aminas ou anéis aromáticos). Não foi possível encontrar um grupo reduzido de descritores com elevada capacidade para prever o ponto de fusão destes sais. O método baseado em redes neuronais CPG permitiu o processamento em simultâneo de quatro famílias aniónicas, produzindo quatro mapas que quando alinhados permitem uma comparação eficaz entre aniões relativamente à relação entre estrutura e ponto de fusão.

Na secção **e)** foram sintetizados novos sais de elevada densidade ficando demonstrado que o método de síntese utilizado, baseado na bromação de posições olefínicas é eficaz, sendo obtidos compostos com elevada pureza, ainda que os rendimentos sejam relativamente baixos. Um dos compostos sintetizados é um líquido iónico, apresentando ponto de fusão de 74,3°C e densidade de 1,68 g/cm³. Em estudo posterior foi possível reduzir o ponto de fusão deste composto na presença de CO₂ super-crítico (44,6 °C na presença de 113 bar de CO₂).

Finalmente no estudo **f)** foram sintetizados cinco novos sais de guanidínio de elevada pureza, baseados em aniões com capacidade de deslocalizar a carga negativa. Os cinco compostos sintetizados são líquidos iónicos, tendo sido obtido um composto líquido à temperatura ambiente baseado no anião *para*-toluenosulfonato. Neste estudo ficou definido que sais contendo aniões com capacidade para deslocalizar carga negativa e/ou organo-fluorados apresentam os pontos de fusão mais reduzidos.

Apesar de na maioria dos estudos os objectivos propostos terem sido cumpridos, é perfeitamente possível melhorar os modelos de previsão, especialmente aqueles que dizem respeito ao ponto de fusão e/ou estado físico já que novos sais foram entretanto sintetizados e caracterizados por diversos grupos permitindo aumentar a variabilidade estrutural dos conjuntos de treino. Tendo em conta que tanto os teores em água e cloretos influenciam significativamente as propriedades físico-químicas de líquidos iónicos cada vez mais líquidos iónicos aparecem caracterizados com estes teores, havendo agora critérios objectivos na selecção de sais a incluir em conjunto de treino.

Tendo em conta a experiência acumulada, em alguns casos seria benéfica uma selecção mais fina de descritores moleculares tendo em conta padrões estruturais presentes em alguns dos compostos com as previsões mais afastadas da realidade. Noutro sentido, seria interessante estudar as propriedades físico-químicas de sais baseados no anião perfluorooctanoato, visto ser um anião pouco conhecido até à data no contexto dos líquidos iónicos tendo permitido a obtenção de um líquido à temperatura ambiente. Seria igualmente muito importante perceber qual a interacção entre CO₂ super-crítico e este tipo de estruturas iónicas.

II-5: Parte Experimental

II-5.1: Preâmbulo

A realização da componente experimental do trabalho apresentado neste capítulo, que consistiu na síntese de novos sais de catião orgânico com propriedades de ponto de fusão e ou densidades pretendidas, contou com uma série de procedimentos gerais descritos em baixo:

- a) Os reagentes utilizados foram adquiridos comercialmente e, salvo indicação em contrário, não foram purificados.
- b) Os solventes utilizados nas sínteses foram previamente destilados na presença de CaH_2 sob atmosfera de árgon.
- c) Para cromatografia em coluna simples (*c.c.*) utilizou-se sílica de granulometria da marca/modelo Macherey-Nagel 60 M (Ref. 815381).
- d) Os solventes utilizados na fase de *work-up* da reacção foram destilados previamente.
- e) Todos os produtos isolados foram submetidos a alto vácuo (10^{-6} mbar) durante *ca* de 5h de forma a eliminar compostos voláteis.
- f) Os Espectros de IV foram traçados num espectrofotómetro Jasco FT/IR-430. Na descrição dos dados experimentais é explicitado, pela seguinte ordem: o **tipo de amostra** utilizada – KBr (pastilha de brometo de potássio para compostos sólidos) ou filme (em células de cloreto de sódio no caso de líquidos e óleos); **frequência** do máximo de uma banda de absorção (ν_{max} em cm^{-1}); **tipo de banda** (quando é relevante para a identificação) – f (forte), fr (fraca) ou l (larga), atribuição a grupo de átomos (quando é possível fazer a atribuição de uma forma inequívoca).
- g) Espectros de Ressonância Magnética Nuclear (RMN) obtidos em espectrómetro Bruker AMX400. Desvios químicos obtidos em ppm (parts per million) usando como referência o tetrametilsilano. Os dados obtidos são indicados com os seguintes parâmetros na seguinte ordem: **solvente**; **desvio químico**; **integração** (no caso de ^1H -RMN); **multiplicidade do sinal** – s (singuleto), sl (singuleto largo), d (duplete), t (triplete), q (quarteto), qt (quintoplete), sx (sexteto), m (multiplete); **constante de acoplamento** (*J*, em Hz para ^1H -RMN); **atribuição na molécula** (quando é possível fazer atribuição de forma inequívoca).

- h) As medidas calorimétricas foram realizadas em sistema 2920MDSC da TA Instruments Inc. Sendo indicadas na caracterização do composto o **ponto de fusão** (*p.f.* – quando existente); **transição vítrea** (*t.v.* – quando existente); **temperatura de decomposição** (*t.d.* - quando obtida). Gás Hélio com caudal de $30 \text{ cm}^3 \cdot \text{min}^{-1}$ atravessou a amostra. O arrefecimento foi obtido através de acessório de arrefecimento por azoto líquido (LNCA), permitindo um decréscimo de temperatura contínuo até $-150 \text{ }^\circ\text{C}$. A linha de base foi calibrada submetendo o porta-amostras vazio à mesma gama de temperaturas usada nos ensaios com amostra. A calibração das temperaturas foi realizada baseada nos picos endotérmicos de fusão associados a diversos padrões: n-decano (*p.f.* = $243,75 \text{ K}$), n-octadecano (*p.f.* = $301,77 \text{ K}$), hexatriacontano (*p.f.* = $347,30 \text{ K}$), Índio (*p.f.* = $430, 61 \text{ K}$) e Estanho (*p.f.* = $506,03 \text{ K}$). A entalpia foi calibrada com Índio ($\Delta H_{\text{fusão}} = 28, 71 \text{ J} \cdot \text{g}^{-1}$).
- i) Para a medição de densidades foram usadas metodologias diferentes consoante a natureza e quantidade da amostra. Para amostras líquidas em volumes superiores a $700 \text{ }\mu\text{l}$ foi utilizado densímetro Anton Paar DMA38 ($0\text{-}3 \text{ g} \cdot \text{cm}^3$) (densidade medida de acordo com este método assinalada com ρ^a). No caso de amostras líquidas em volumes inferiores a $700 \text{ }\mu\text{l}$ foi utilizado micropicnómetro:



Figura II.40: Micropicnómetro utilizado para medir densidade de líquidos em volumes inferiores a $700 \text{ }\mu\text{l}$.

Este picnómetro foi calibrado com parafina, possuindo um volume de $214 \text{ }\mu\text{l}$ (a 22°C), sendo validado comparando medição entre micropicnómetro e densímetro Anton Paar DMA38 para poliglicol B01/240 ($\text{CH}_3\text{CH}_2\text{CH}_2\text{CH}_2(\text{OCH}_2\text{CHCH}_3)_m\text{OH}$ massa molar aproximadamente $2300 \text{ g} \cdot \text{mol}^{-1}$) (densidade medida de acordo com este método assinalada com ρ^b). Por sua vez a densidade de amostras sólidas foi medida por método indirecto usando picnómetro artesanal:



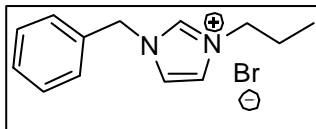
Figura II.41: Picnómetro usado para medir densidade de sólidos.

Este picnómetro foi calibrado com decano a 25°C (1,085 mL) e validado para a determinação de densidade de cloreto de sódio cuja densidade existe tabelada; nestas medições sabendo a massa de sólido introduzida no picnómetro, a massa de líquido que não dissolvesse o sólido (decano) utilizado para perfazer o volume do picnómetro e sabendo ainda a densidade do decano à temperatura das medições (25°C) era possível saber a densidade do sólido, (densidade medida de acordo com este método assinalada com ρ°).

- j) Os espectros de massa foram obtidos na unidade de espectrometria de massa da Universidade de Santiago de Compostela, Equipamento Micromass, modelo Autospec, a técnica de ionização utilizada foi FAB (Fast Atom Bombardment) e IE (Impacto Electrónico). Adicionalmente foi usado o modelo micrOTOF da Bruker e a técnica de ionização utilizada foi ESI (Electrospray Ionization). Os dados obtidos são indicados na seguinte ordem: **razão massa carga** (m/z); **atribuição** a ião ou outro fragmento molecular; **intensidade dos picos** em relação ao pico mais intenso (100%). Na interpretação dos espectros M representa a massa molecular. Os dados dos espectros de massa de alta resolução (EMAR) são indicados pela seguinte ordem: **razão massa carga** do ião molecular; **fórmula molecular** e **massa exacta** teórica do ião molecular correspondente
- k) As microanálises elementares foram realizadas em equipamento (Thermo Finnigan-CE Instruments, modelo Elemental Analyser 1112 series). Do Departamento de Química da Faculdade de Ciências e tecnologia da Universidade Nova de Lisboa. Os dados obtidos estão na seguinte ordem: **fórmula molecular**; **percentagem calculada** para elementos do composto analisado e **percentagem determinada** para elementos desse composto.
- l) A titulação da quantidade de água foi feita em aparelho Metrohm KF Coulometer 831 equipado com 703 TI stand e 768 KF oven. As amostras foram tituladas num forno a 130°C com um fluxo de ar 100mL/min.

II-5.2: Síntese de Sais de Imidazólio indicados por Modelo de Previsão de Pontos de Fusão

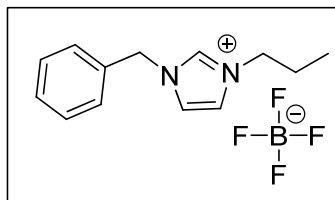
II-5.2.1: Brometo de 1-benzil-3-propil imidazólio (1a)



A uma solução de benzilimidazólio (1 g, 6,32 mmol) em acetonitrilo (3 mL) foi adicionado 1-bromopropano (2,02 g, 16,4 mmol). A mistura resultante ficou sob agitação durante 24 h à temperatura ambiente. Na fase de *work-up* o solvente foi evaporado e a mistura resultante foi lavada com éter etílico. O produto resultante apresenta-se na forma de um líquido viscoso de cor laranja (1,496 g, 84,5%). *t.v.* -38,32°C, *t.d.* 250°C. **IV_{vmax} (filme):** 3130 (CH_{aromática}), 3066 (CH_{aromática}), 2968 (CH_{alifática}), 2877 (CH_{alifática}), 1626 (C-N), 1560 (anel imidazólio), 1498 (CC_{aromática}), 1456 (CC_{aromática}), 1406 (CC_{aromática}), 1387, 1361, 1325, 1207, 1159 (2-CH), 1111, 1082, 1030, 864, 823, 800 cm⁻¹. **¹H-RMN (400 MHz, CDCl₃)** δ: 0,92-0,96 (3H, t, J = 7,6 Hz, CCH₃), 1,90-1,96 (2H, sx, J₁ = 7,2, J₂ = 7,6 Hz, CCH₂C), 4,23-4,27 (2H, t, J = 7,2 Hz, NCH₂C_{alifático}), 5,59 (2H, s, NCH₂C_{aromático}), 7,48-7,26 (7H, m, CH_{aromático} + CH_{olefínico}), 10,53 (1H, s, NCHN). **¹³C-RMN (100 MHz, CDCl₃)** δ: 10,74 (CCH₃), 23,60 (CH₂CH₃), 51,61 (NCH₂C_{alifático}), 53,26 (NCH₂C_{aromático}), 121,97 e 122,28 (C_{olefínico}), 129,35 (C_{aromático}), 129,39 (C_{aromático}), 129,43 (C_{aromático}), 133,12 (C_{aromático}), 136,92 (NCHN). **Análise elementar calculada para C₁₃H₁₇N₂Br.0.7H₂O (%):** N 9,53, C 53,14, H 6,31; **composição determinada (%):** N 9,69, C 53,59, H 6,17.

II-5.2.2: Compostos Resultantes da troca Aniônica Realizada Partindo de Brometo de 1-benzil-3-propil imidazólio (1a)

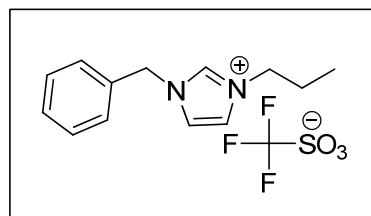
II-5.2.2.1: Tetrafluoroborato de 1-benzil-3-propil imidazólio (2a)



A uma solução de brometo de 1-benzil-3-propil imidazólio (**1a**) (0,25 g, 0,89 mmol) em 10 mL de diclorometano foi adicionado tetrafluoroborato sódico (0,127g, 1,16 mmol). A mistura resultante ficou em agitação contínua durante 24 h à temperatura ambiente. Na fase de *work-up* a mistura foi filtrada e posteriormente eluída em coluna cromatográfica com empacotamento de sílica-

gel (*c.c.*), o solvente de eluição foi o diclorometano. Depois de evaporar solvente sob pressão reduzida foi obtido um líquido viscoso (0,206 g, 80%). **IV_{vmax} (filme):** 3149 (CH_{aromática}), 3089 (CH_{aromática}), 2970 (CH_{alifática}), 2879 (CH_{alifática}), 1626 (C-N), 1587, 1562, 1498 (CC_{aromática}), 1458 (CC_{aromática}), 1406 (CC_{aromática}), 1388, 1360, 1331, 1286, 1207, 1159 (2-CH), 1061 (BF₄⁻), 904, 856, 823, 800 cm⁻¹. **¹H-RMN (400 MHz, CDCl₃)** δ: 0,88-0,92 (3H, t, J = 7,6 Hz, CCH₃), 1,87-1,91 (2H, sx, J₁ = 7,2, J₂ = 7,6 Hz, CCH₂C), 4,15-4,19 (2H, t, J = 7,2 Hz, NCH₂C_{alifático}), 5,45 (2H, s, NCH₂C_{aromático}), 7,26-7,43 (7H, m, CH_{aromático} + CH_{olefínico}), 9,71 (1H, s, NCHN). **¹³C-RMN (100 MHz, CDCl₃)** δ: 13,42 (CCH₃), 19,42 (CCH₂C), 32,17 (NCH₂C), 50,09 (NCH₂C_{aromático}), 121,58, 122,91 (C_{olefínico}), 129,97 (C_{aromático}), 130,45 (C_{aromático}), 130,60 (C_{aromático}), 134,48 (C_{aromático}), 136,28 (C_{aromático}), 136,46 (NCH₂N). **Análise elementar calculada para C₁₃H₁₇N₂BF₄·0,3H₂O (%):** N: 9,54, C: 53,20, H 6,04. **composição determinada (%):** N 9,35, C 53,29, H 6,62.

II-5.2.2.2: Trifluorometanosulfonato de 1-benzil-3-propil imidazólio (3a)



Numa solução de brometo de 1-benzil-3-propil imidazólio (**1a**) (0,25 g, 0,89 mmol) em diclorometano (10 mL) foi adicionado trifluorometanosulfonato de potássio (0,218 g, 1,16 mmol). A mistura ficou em agitação contínua durante 24 h à temperatura ambiente. Na fase de *work-up* a mistura foi filtrada e eluída numa coluna cromatográfica (*c.c.*) com empacotamento de sílica-gel, o solvente de eluição foi o diclorometano. O produto resultante foi obtido na forma de cristais brancos (0,225 g, 72,2%). *p. f.* 55,03°C. **IV_{vmax} (KBr):** 3157 (CH_{aromática}), 3095 (CH_{aromática}), 2972 (CH_{alifática}), 2883 (CH_{alifática}), 2293, 1979, 1923, 1846, 1701, 1593, 1562, 1498 (CC_{aromática}), 1456 (CC_{aromática}), 1408 (CC_{aromática}), 1390, 1350, 1263 (CF₃SO₃⁻), 1225, 1163 (2-CH), 1090, 1032, 937, 903, 856, 829, 800, 781, 756, 739, 719, 636, 573, 517 cm⁻¹. **¹H-RMN (400 MHz, CDCl₃)** δ: 0,90-0,93 (3H, t, J = 7,6 Hz), 1,83-1,92 (2H, sx, J₁ = 7,6, J₂ = 7,6 Hz, CCH₂C), 4,12-4,15 (2H, t, J = 7,6 Hz, NCH₂C_{alifático}), 5,35 (2H, s, NCH₂C_{aromático}), 7,26-7,39 (7H, m, CH_{aromático} + CH_{olefínico}), 9,27 (1H, s, NCHN). **¹³C-RMN (100 MHz, CDCl₃)** δ: 10,48 (CCH₃), 23,34 (CCH₂C), 51,57 (NCH₂C), 53,39 (NCH₂C_{aromático}), 122,09, 122,25, 122,43 (C_{olefínico}), 128,85 (C_{aromático}), 129,45 (C_{aromático}), 129,52 (C_{aromático}), 132,84 (C_{aromático}), 136,08 (NCH₂N) ppm. **Análise elementar calculada para C₁₄H₁₇N₂F₃O₃S (%):** N: 8,00 C: 47,99, H 4,89. **composição determinada (%):** N 7,81, C 47,94, H 5,15.

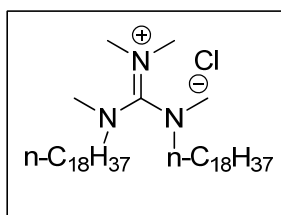
II-5.3: Síntese de Sais de Guanidínio Indicados por Modelo de Previsão de Pontos de Fusão

II-5.3.1: Síntese de cloretos de Guanidínio

II-5.3.1.1: Método Geral

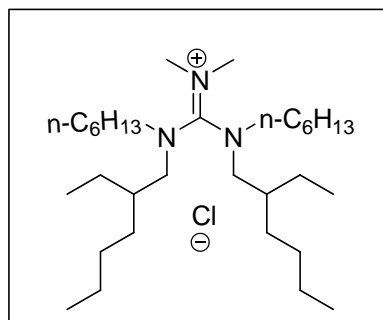
Todos os cloretos de guanidínio foram preparados de acordo com procedimento publicado:⁴² A uma mistura sob agitação de cloreto de N,N-dimetil-fosgenimínio (3,4 mmol) em diclorometano anidro (10 mL) a 0 °C (banho de gelo) e sob atmosfera de árgon, foi adicionada muito lentamente uma mistura de amina secundária (2,1 eq.) e trietilamina anidra (2,01 mL, 2,2 eq.) dissolvidas em diclorometano anidro (14 mL). Após adição a mistura reaccional foi mantida à temperatura de 0°C durante 30 minutos. Findo esse tempo mistura reaccional permaneceu sob agitação à temperatura ambiente. Para cada caso são especificadas as quantidades, tempo de reacção e procedimento de *work-up* utilizado.

II-5.3.1.1.1: Cloreto de N,N,N',N''-tetrametil-N',N''-dioctadecilguanidínio (4a)



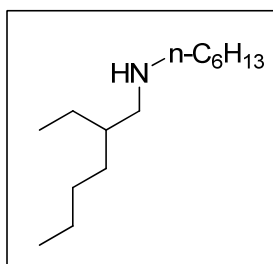
Sintetizado de acordo com o método geral usando 0,456 g (2,81 mmol) de cloreto de fosgenimínio, 2 g (7,05 mmol) de metiloctadecilamina e 0,62 g (6,81 mmol) de trietilamina em 7 mL de diclorometano no total, a fase reaccional durou 20 horas. Na fase de *work-up* a mistura foi filtrada e o sólido obtido lavado com diclorometano. Posteriormente o solvente das fases orgânicas combinadas foi removido a pressão reduzida e a amostra resultante foi adicionada água (12 mL). A fase aquosa foi extraída com éter etílico (2 × 10 mL). As fases orgânicas combinadas foram lavadas com solução aquosa de ácido clorídrico (1 M). O solvente foi removido sob pressão reduzida, sendo obtido sal de guanidínio apresentando-se na forma de cristais brancos (0,93 g, 40%). *p.f.* 78°C, *t.d.* 250°C **IV_{vmax} (KBr)** 3248, 2920 (f, CH_{alifática}), 2850 (f, CH_{alifática}), 1593 (C-N), 1566 (C-N), 1469, 1412, 1377, 1304, 1254, 1207, 1157, 1138, 1095, 1068, 903, 721 cm⁻¹. **¹H-RMN (400 MHz, CDCl₃)** δ: 0,85 (6H, t, J = 6,8 Hz, CCH₃), 1,22 (64H, m, CCH₂C), 2,92-3,10 (16H, m, NCH₃, NCH₂C); **¹³C-RMN (100 MHz, CDCl₃)** δ: 14,03 (CCH₃), 22,61 (CCH₂C), 26,76 (CCH₂C), 27,52 (CCH₂C), 29,28 (CCH₂C), 31,85 (CCH₂C), 38,62 e 38,86 (NCH₃, rotâmeros), 40,38 e 41,36 (NCH₃, rotâmeros), 52,80 (NCH₂C) e 163,73 (C(NR₁R₂)₂(NRR)); **Análise elementar calculada para C₄₁H₈₆ClN₃·1,5H₂O (%):** N 6,15, C 72,03, H 13,12; **composição determinada (%):** N 6,24, C 72,06, H 13,45.

II-5.3.1.1.2: Cloreto de *N,N*-dimetil-*N',N''*-dihexil-*N',N''*-bis(2-etilhexil)guanidínio (5a)



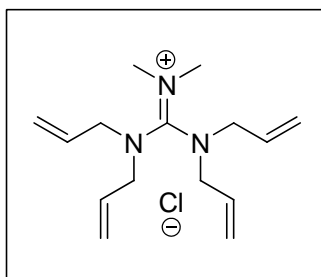
Sintetizado de acordo com o método geral usando (2-etil-hexil)-hexilamina (1,95 g, 9,1 mmol), preparada de acordo com procedimento descrito no tópico seguinte, cloreto de fosgeniminio (0,705 g, 4,34 mmol) e 1,70 mL de trietilamina (9,46 mmol) no total de 21 mL diclorometano, a fase reaccional durou 6 horas. Na fase de *work-up* a mistura foi filtrada e o sólido foi lavado com diclorometano. O solvente das fases orgânicas combinadas foi removido a pressão reduzida e ao resíduo obtido foi adicionada água (15 mL). A fase aquosa foi extraída três vezes com éter etílico (3x25 mL). As fases orgânicas combinadas foram lavadas com solução aquosa de ácido clorídrico (1 M). Posteriormente o solvente foi removido a pressão reduzida com o sal de guanidínio a ser obtido como um óleo amarelo (1,419 g, 64%). *t.v.* -85,94 °C. **IV_{vmax} (filme):** 2927 (f, CH_{alifática}), 2857 (f, CH_{alifática}), 1643 (C-N), 1577 (C-N), 1535 (C-N), 1462, 1377, 1246, 1188, 1138, 1061, 879, 783, 729 cm⁻¹. **¹H-RMN (400 MHz, CDCl₃)** δ: 0,80-0,90 (18H, m, CCH₃), 1,16-1,27 (28H, m, CCH₂C), 1,41-1,53 (4H, m, CHCH₂CH₃), 1,93-1,96 (2H, m, CH(R₁R₂R₃)), 2,76 (6H, s, NCH₃), 3,05 (4H, d, J=7,2 Hz, N-CH₂C), 3,17 (4H, m, N-CH₂C). **¹³C-RMN (100 MHz, CDCl₃)** δ: 10,04 e 10,45 (CCH₃, rotâmeros), 13,83, 13,91 e 13,96 (CCH₃, rotâmeros), 22,41 e 22,78 (CCH₂CH₃, rotâmeros), 23,58 e 23,71 (CCH₂CH₃, rotâmeros), 26,41 e 26,55 (CCH₂C, rotâmeros), 27,61, 28,14 e 28,53 (CCH₂C, rotâmeros), 30,31 e 30,46 (CCH₂C, rotâmeros), 31,14 e 31,45 (CCH₂C, rotâmeros), 35,88 (CH(CH₂)₃), 38,70 (N-CH₃), 48,21 e 48,68 (N-CH₂C, rotâmeros), 50,69 e 51,05 (N-CH₂C, rotâmeros), 165,17 (C(NR₁R₂)₂(NRR)). **Análise elementar calculada para C₃₁H₆₆ClN₃·1,25H₂O (%):** N 7,80 C 69,10 H 12,81. **composição determinada (%):** N 7,52 C 68,96 H 13,35.

II-5.3.1.1.3: (2-Etil-hexil)-hexilamina (5b)



Preparada em dois passos. A amina primária (2-Etil-hexil)-amina (1 g, 7,8 mmol) foi dissolvida em diclorometano anidro (20 mL) na presença de peneiros moleculares 4 Å (4 g) sendo adicionado hexanal (1,1 g, 10,78 mmol) à temperatura ambiente. A mistura ficou sob agitação durante 24 h sob atmosfera de árgon. Posteriormente a mistura foi filtrada e o solvente da fase orgânica recolhida foi removido a pressão reduzida. O resíduo resultante foi dissolvido éter etílico anidro (30 mL) e LiAlH_4 (0,61 g, 16 mmol) foi adicionado a 0°C (banho de gelo). A mistura permaneceu em agitação contínua por 24 h à temperatura ambiente sob atmosfera de árgon. Na fase de *work-up*, o LiAlH_4 foi neutralizado através adição lenta de água à mistura reaccional, colocada à temperatura de 0°C (banho de gelo). A adição de água cessou quando se formou um precipitado branco consolidado no fundo do balão. A mistura reaccional foi filtrada. Posteriormente a mistura resultante foi destilada a pressão reduzida com a amina secundária a ser recolhida a 66°C (0,3 mmHg) como um líquido transparente (1g, 60%). $^1\text{H-RMN}$ (400 MHz, CDCl_3) δ : 0,83-0,89 (9H, m, CCH_3), 1,26-1,34 (14H, m, CCH_2C), 1,40-1,49 (3H, m, $(\text{CH}_2)_3\text{CH}$, CCH_2C), 2,47-2,48 (2H, d, $J = 6,2$ Hz, $\text{NCH}_2\text{CH-}$), 2,55-2,58 (2H, t, $J = 7,28$ Hz, $\text{NCH}_2\text{CH}_2\text{-}$). $^{13}\text{C-RMN}$ (100 MHz, CDCl_3) δ : 10,83 (CCH_3), 14,02 (CCH_3), 14,09 (CCH_3), 22,63 (CCH_2CH_3), 23,11 (CCH_2CH_3), 24,52 (CCH_2CH_3), 27,09 (CCH_2C), 29,01 (CCH_2C), 30,00 (CCH_2C), 31,80 (CCH_2C), 32,86 (CCH_2C), 39,34 ($\text{CH}(\text{CH}_2)_3$), 50,36 (NCH_2C), 53,29 (NCH_2C).

II-5.3.1.1.4: Cloreto de *N,N*-dimetil-*N',N',N'',N''*-tetraalilguanidínio (6a)



Preparado de acordo com procedimento geral a partir da dialilamina (2 g, 20,6 mmol), cloreto de fosgenimínio (1,59g, 9,8 mmol) e 3,84 mL (21,56 mmol) de trietilamina em 47 mL de diclorometano no total, a fase reaccional durou 10 horas. Na fase de *work-up* a mistura foi filtrada e sólido lavado com diclorometano. O solvente das fases orgânicas foi removido sob pressão reduzida

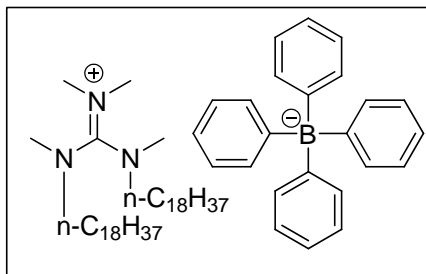
com solução aquosa de NaOH (2 M; 16 mL) a ser adicionada à mistura. A fase aquosa foi lavada duas vezes com éter dietílico (2 × 22 mL) e o pH foi ajustado a 7 por adição lenta de ácido clorídrico (37%). A água foi removida sob pressão reduzida e o sal de guanidínio foi obtido como um sólido castanho pálido (1,8 g, 65%). *p.f.* 79,2°C; **IV_{max} (KBr)**: 3082 (CH_{olefinica}), 2978 (CH_{alifática}), 2920 (CH_{alifática}), 1643 (C-N), 1589 (C-N), 1539 (C-N), 1419, 1338, 1277, 1234, 1146, 1061, 999 (CH_{olefinica}), 933 (CH_{olefinica}), 876 (CH_{olefinica}), 671 cm⁻¹. **¹H-RMN (400 MHz, CDCl₃)** δ: 3,06 (6H, s, NCH₃), 3,78-3,96 (8H, m, NCH₂CH-), 5,28-5,32 (8H, m, CH=CH₂), 5,76 (4H, m, CH=CH₂). **¹³C-RMN (100 MHz, CDCl₃)** δ: 40,97 (NCH₃), 52,64 e 52,89 (NCH₂CH=CH₂, rotâmeros), 121,57 (CH=CH₂), 131,11 (CH=CH₂), 163,03 (C(NR₁R₂)₂(NRR)). **Análise elementar calculada para C₁₅H₂₆ClN₃·1,7H₂O (%)**: N 13,36, C 57,29, H 9,42. **composição determinada (%)**: N 13,48, C 56,91, H 9,63. **m/z (IE⁺)** 248, (M⁺, 100), 206 (C₁₂H₂₀N₃⁺, 8,15), 192 (C₁₁H₁₈N₃⁺, 11,69), 166 (C₉H₁₆N₃⁺, 30,91), 111 (C₆H₁₁N₂⁺, 89,15), 71 (C₃H₇N₂⁺, 10,19), **EMAR**: 248,212524 (C₁₅H₂₆N₃⁺ requer 248,212673).

II-5.3.2: Síntese de Sais de tetrafenilborato de Guanidínio

II-5.3.2.1: Método Geral

A partir de cloreto de guanidínio preparado de acordo com o procedimento geral do tópico II-5.3.1.1. dissolvido em diclorometano, sendo adicionado tetrafenilborato sódico (número de equivalentes específico em cada caso) e agitado por 24 h O sólido resultante foi recolhido por filtração e lavado com diclorometano; as fases orgânicas combinadas foram secas com MgSO₄, filtradas, parcialmente evaporadas e eluídas com diclorometano numa coluna de sílica gel e carvão activado. O solvente foi removido sob pressão reduzida sendo obtido o respectivo tetrafenilborato de guanidínio. São especificadas as quantidades e o procedimento de *work-up* utilizado.

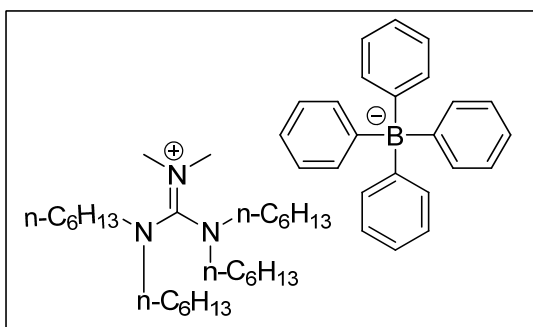
II-5.3.2.1.1: Tetrafenilborato de *N, N, N', N''*-Tetrametil-*N', N''*-dioctadecilguanidínio (7a)



Preparado a partir de cloreto de *N, N, N', N''*- Tetrametil-*N', N''*-dioctadecil guanidínio (**4a** – preparado segundo descrição do tópico II-5.3.1.1.1) (0,2 g, 0,29 mmol) em 15 mL de diclorometano de acordo com procedimento geral para substituição de anião cloreto por tetrafenilborato (0,184 g, 0,585 mmol). Na fase de *work-up* a mistura foi filtrada e o sólido obtido lavado com diclorometano.

As fases orgânicas combinadas foram concentradas e eluídas com diclorometano através de uma coluna de sílica gel e carvão activado. O solvente foi removido sob pressão reduzida. O sal de guanidínio foi obtido como um sólido branco (0,164 g, 59,7%). *p.f.* 88°C; *t.d.* > 300°C; **IV_{vmax} (KBr)**: 3051 (fr), 3051 (CH_{aromática}), 2920 (CH_{alifática}), 2850 (CH_{alifática}), 1566 (C-N), 1466 e 1408 (C=C_{aromática}), 1265, 1149, 1061, 1030, 895, 849, 741, 706, 606 cm⁻¹. **¹H-RMN (400 MHz, CDCl₃)** δ: 0,89 (6H, t, J = 6,92 Hz, CCH₃), 1,27 (64H, m, CCH₂C), 2,48 (12H, m, NCH₃), 2,81 (4H, m, NCH₂C), 6,89 (4H, t, J = 6,92 Hz, CH_{aromático,anião}), 7,03 (8H, t, J = 7,2 Hz, CH_{aromático,anião}), 7,42 (8H, m, CH_{aromático,anião}). **¹³C-RMN (100 MHz, CDCl₃)** δ: 14,07 (CCH₃), 22,65 (CCH₂CH₃), 26,68 (CCH₂C), 27,47 (CCH₂C), 29,17, 29,32, 29,55, 29,67 (CCH₂C), 37,68 (NCH₃), 39,91 e 40,20 (NCH₃, rotâmeros), 52,62 (NCH₂C), 121,69, 125,51, 136,17 (C_{aromático}), 163,00 (C(NR₁R₂)₂(NRR)), 163,34 163,83, 164,32, 164,81 (C_{aromático}B). **Análise elementar calculada para C₆₅H₁₀₆BN₃ (%)**: N 4,47, C 83,02, H 11,36. **composição determinada (%)**: N 4,53, C 82,73, H 11,63.

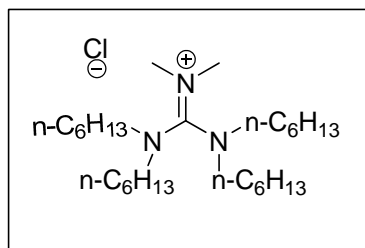
II-5.3.2.1.2: Tetrafenilborato de *N,N*-dimetil-*N',N',N'',N''*-tetrahexilguanidínio (8a)



Preparado a partir de cloreto de *N,N*-dimetil-*N',N',N'',N''*-tetrahexilguanidínio, previamente sintetizado (**8b**) de acordo com procedimento descrito na literatura para a sua síntese⁴² (1g, 2,17 mmol), e tetrafenilborato sódico (1,39g, 4,37 mmol) em 20 mL de diclorometano de acordo com procedimento geral de substituição de anião cloreto por tetrafenilborato. Na fase de *work-up*, a mistura foi filtrada e o sólido obtido foi lavado com diclorometano. As fases orgânicas combinadas foram concentradas e eluídas com diclorometano através de coluna de sílica gel e carvão activado (*c.c.*). O solvente foi removido sob pressão reduzida sendo obtido um sólido castanho (1g, 62%). *p.f.* 58°C. **IV_{vmax} (KBr)**: 3055 (CH_{aromática}), 2927 e 2858 (CH_{alifática}), 1574 (C-N), 1543 (C-N), 1462 (C_{aromática}), 1377, 1265, 1142, 1034, 845, 706, 613 cm⁻¹. **¹H-RMN (400 MHz, CDCl₃)** δ: 0,83-0,92 (12H, t, J = 6,8 Hz, CCH₃), 1,19-1,32 (32H, m, CCH₂C, CCH₂CH₃), 2,21 (6H, s, NCH₃), 2,63-2,97 (8H, m), 6,91 (4H, t, J = 7,16 Hz, CH_{aromático}), 7,06 (8H, t, J = 7,2 Hz, CH_{aromático}), 7,48 (8H, m, CH_{aromático}). **¹³C-RMN (100 MHz, CDCl₃)** δ: 13,88 (CCH₃), 22,32 (CCH₂C), 26,62 (CCH₂C), 27,32 (CCH₂C), 27,83 (CCH₂C), 31,25 (CCH₂C) e 31,33 (CCH₂C), 39,78 (NCH₃), 49,29 (NCH₂C), 50,10 (NCH₂C), 121,71 (C_{aromático}), 125,64 (C_{aromático}), 136,12 (C_{aromático}), 162,79 (C(NR₁R₂)₂(NRR)), 163,40 (C_{aromático}B), 163,89 (C_{aromático}B), 164,38 (C_{aromático}B), 164,87 (C_{aromático}B). **Análise elementar**

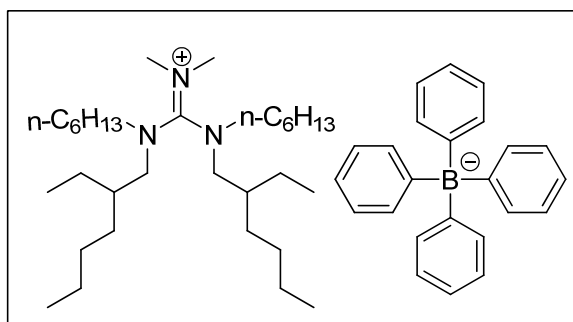
calculada para $C_{51}H_{78}BN_3 \cdot 0,1H_2O$ (%): N 5,63, C 82,13, H 10,57. **composição determinada (%)**: N 5,59, C 82,00, H 10,68.

II-5.3.2.1.3: Cloreto de *N,N*-dimetil-*N',N'',N'''*-tetrahexilguanidínio (8b)



Obtido de acordo com procedimento descrito na literatura para a sua síntese.⁴² **¹H-RMN (400 MHz, CDCl₃)** δ : 0,69-0,73 (12H, m, CCH₃), 1,14 (32H, sl, CCH₂C), 2,98 (6H, s, NCH₃), 3,08-3,12 (8H, t, J = 7,12, NCH₂C). **¹³C-RMN (100 MHz, CDCl₃)** δ : 13,53 e 13,58 (CCH₃, rotâmeros), 22,11 (CCH₂C), 26,10, 26,18, 26,28 (CCH₂C, rotâmeros), 27,20, 27,54 (CCH₂C), 39,95, 31,08 (CCH₂C, rotâmeros), 40,61 (NCH₃), 49,45 (NCH₂C), 164,86 (C(NR₁R₂)₂(NRR)).

II-5.3.2.1.4: Tetrafenilborato de *N,N*-Dimetil-*N',N''*-dihexil-*N',N''*-bis(2-etilhexil)guanidínio (9a)



Preparado a partir de cloreto de *N,N*-dimetil-*N',N''*-dihexil-*N',N''*-bis(2-etilhexil)guanidínio (**5a**) (0,5 g, 0,97 mmol) em 11 mL de diclorometano de acordo com procedimento geral para a substituição de anião cloreto por tetrafenilborato (1,495 g, 2,9 mmol). Na fase de *work-up*, a mistura foi filtrada e o sólido obtido lavado com diclorometano. As fases orgânicas combinadas foram concentradas e posteriormente eluídas com diclorometano através de uma coluna de sílica carvão activado (*c.c.*). O sal de guanidínio foi obtido na forma de um sólido castanho (0,63 g, 82%). *p.f.* 125,6°C, **IV_{vmax} (filme)** 3157 (CH_{aromática}), 3055 (CH_{aromática}), 2929 e 2858 (CH_{alifática}), 1579 (C-N), 1537 (C-N), 1502 (C-N), 1464, 1404, 1379, 1269, 1244, 1184, 1146, 1066, 1032, 868, 847, 735, 706 (CH_{aromática}), 665, 611 cm⁻¹. **¹H-RMN (400 MHz, CDCl₃)** δ : 0,75-0,84 (18H, m, CCH₃), 1,15-1,21 (28H, m, CCH₂C), 1,33-1,42 (4H, m, CCH₂C), 1,81-1,86 (2H, m, CH(CH₂)₃), 2,71 (6H, s, NCH₃), 2,96 (4H, d, J = 7,24 Hz, NCH₂C), 3,03-3,25 (4H, m, NCH₂C), 7,27-7,29 (4H, m, CH_{aromático}), 7,35-

7,39 (8H, m, $\text{CH}_{\text{aromático}}$), 7,73-7,79 (8H, m, $\text{CH}_{\text{aromático}}$). ^{13}C -RMN (100 MHz, CDCl_3) δ : 10,04 e 10,46 (CH_2CH_3 , rotâmeros), 13,97 (CH_2CH_3 , rotâmeros), 22,43, 22,53, 23,56 e 23,71 (CH_2CH_3 , rotâmeros), 26,46, 26,52 (CCH_2C , rotâmeros), 27,66, 28,15 e 28,56 (CCH_2C , rotâmeros), 30,29, 30,48, 31,15 e 31,50 (CCH_2C , rotâmeros), 35,92 ($\text{CH}(\text{CH}_2)_3$), 37,14, 38,58 (NCH_3), 48,19, 48,51, 50,69 e 50,88 (NCH_2C , rotâmeros), 115,46, 119,81, 121,60, 125,47, 127,09, 127,18, 127,40, 127,81, 128,26, 128,68, 129,33, 130,26, 130,90, 134,10, 134,70, 135,58 e 136,23 ($\text{CH}_{\text{aromático}}$), 165,87 ($\text{C}(\text{NR}_1\text{R}_2)_2(\text{NRR})$). **Análise elementar calculada para $\text{C}_{55}\text{H}_{86}\text{BN}_3 \cdot 0,9\text{H}_2\text{O}$ (%):** N 5,20, C 80,87, H 10,80. **composição determinada (%):** N 4,58, C 80,51, H 10,52.

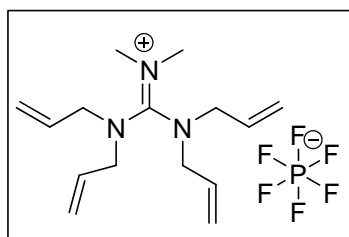
II-5.4: Síntese de Líquidos Iônicos Densos, Sais de Guanidínio:

Os compostos caracterizados neste tópico foram obtidos com a intenção de obter compostos de elevada densidade, alguns destes compostos foram obtidos em até 4 passos reacionais, estando caracterizados esses produtos finais e intermediários que lhes deram origem. Este trabalho pode-se dividir em três tópicos principais baseando-se cada um no respectivo cloreto de guanidínio precursor.

II-5.4.1: Compostos Derivados do cloreto de *N,N*-dimetil-*N',N',N'',N''*-tetraalilguanidínio (6a)

Preparado cloreto de *N,N*-dimetil-*N',N',N'',N''*-tetraalilguanidínio (3a) de acordo com procedimento descrito no tópico 5.3.1.1.3 a partir da dialilamina e cloreto de fosgenimínio.

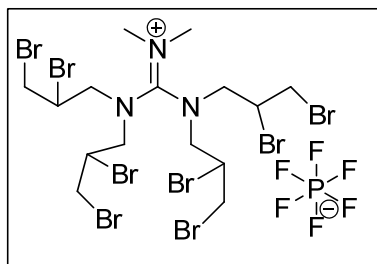
II-5.4.1.1: Hexafluorofosfato de *N,N*-dimetil-*N',N',N'',N''*-tetraalilguanidínio (10a)



Preparado através da adição de hexafluorofosfato de potássio (0,407 g, 2,20 mmol) a solução de cloreto de *N,N*-dimetil-*N',N',N'',N''*-tetraalilguanidínio (6a) (0,5 g, 1,76 mmol) em 25 mL de diclorometano. A mistura resultante ficou em agitação contínua à temperatura ambiente durante 24 h. Na fase de *work-up* a mistura foi filtrada e evaporado o solvente da fase orgânica líquida. No final foi obtido um sólido laranja (0,62 g, 89%). *p. f.* 56,09°C, *t. d.* 232°C, ρ^c 1,21 g/cm³ (25°C). **IV_{vmax} (KBr):** 3088 ($\text{CH}_{\text{olefínica}}$), 3032 ($\text{CH}_{\text{olefínica}}$), 2988 ($\text{CH}_{\text{alifática}}$), 2945 ($\text{CH}_{\text{alifática}}$), 2912 ($\text{CH}_{\text{alifática}}$), 1645 (C-N), 1583 (C-N), 1535 (C-N), 1448, 1417, 1356, 1339, 1278, 1267, 1236, 1159, 1112, 1094, 1067, 1002 ($\text{CH}_{\text{olefínica}}$), 956 ($\text{CH}_{\text{olefínica}}$), 929 ($\text{CH}_{\text{olefínica}}$), 837 (P-F), 678, 648, 557 cm⁻¹. **^1H -RMN (400 MHz, CDCl_3) δ :** 2,87 (6H, s, NCH_3), 3,68 (8H, sl, $\text{NCH}_2\text{CH}=\text{CH}_2$), 5,25-5,29 (8H, m, $-\text{CH}_2\text{CH}=\text{CH}_2$), 5,66

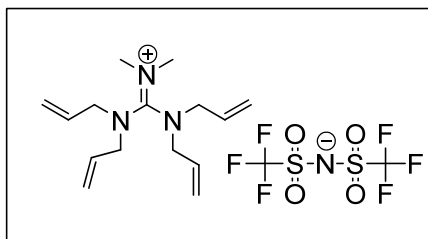
(4H, sl, $-\text{CH}_2\text{CH}=\text{CH}_2$) ppm. ^{13}C -RMN (100 MHz, CDCl_3) δ : 39,94 (NCH_3 , rotâmeros), 52,03 e 52,49 ($\text{NCH}_2\text{CH}=\text{CH}_2$, rotâmeros), 121,25 e 121,46 ($\text{NCH}_2\text{CH}=\text{CH}_2$, rotâmeros) 130,63 ($\text{NCH}_2\text{CH}=\text{CH}_2$), 162,69 ($\text{C}(\text{NR}_1\text{R}_2)_2(\text{NRR})$) ppm. **Análise elementar calculada para $\text{C}_{15}\text{H}_{26}\text{N}_3\text{F}_6\text{P}$ (%):** N: 10,68 C: 45,80, H 6,66. **composição determinada (%):** N 10,62, C 45,96, H 6,60.

II-5.4.1.2: Hexafluorofosfato de N,N -dimetil- N',N',N'',N'' -tetraquis(2,3-dibromopropil) guanidínio (11a)

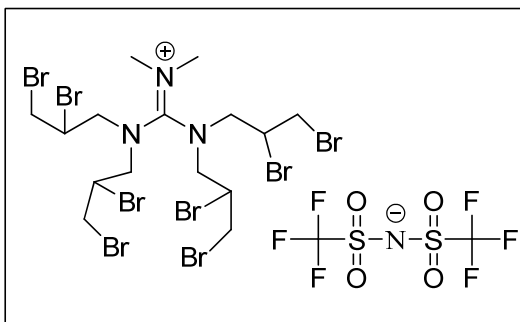


Adaptando procedimento descrito na literatura⁹² este composto foi preparado a partir de hexafluorofosfato de N,N -dimetil- N',N',N'',N'' -tetraalilguanidínio (**10a**) (0,4 g, 1,02 mmol) dissolvido em diclorometano (20 mL), a essa solução foi adicionado bromo (Br_2 - 0,835g, 5,29 mmol). A mistura resultante ficou em agitação contínua em refluxo durante 24h, a reação foi seguida por cromatografia de camada fina *c.c.f.*. Findo esse período procedeu-se ao *work-up* da reação, fase em que a mistura foi filtrada e a fase orgânica líquida recolhida foi lavada 2x10 mL de solução aquosa de sulfito de sódio (10%), posteriormente a mistura foi lavada com 3x10 mL de água. Seguidamente procedeu-se à cristalização do produto com mistura diclorometano/éter dietílico, sendo adicionado éter etílico até o produto começar a cristalizar. No final foi obtido um sólido branco (0,144 g, 13,8%). *p. f./t. d.* 137,5°C ρ^c 1,55 (25°C). **IV_{vmax} (KBr):** 3001 ($\text{CH}_{\text{alifática}}$), 2935 ($\text{CH}_{\text{alifática}}$), 2362, 2345, 1585 (C-N), 1517 (C-N), 1419, 1383, 1313, 1284, 1242, 1151, 1124, 1084, 1063, 949, 935, 901, 840 (P-F), 655 (C-Br), 575 (C-Br) cm^{-1} . **^1H -RMN (400 MHz, CDCl_3)** δ : 3,30-3,38 (8H, m, NCH_2CBr), 3,79-3,88 (14H, m, NCH_3 + $\text{NCH}_2\text{CHBrCH}_2\text{Br}$), 4,04-4,16 (4H, m, $\text{NCH}_2\text{CHBrCH}_2\text{Br}$) ppm. **^{13}C -RMN (100 MHz, CDCl_3)** δ : 33,89, 33,94, 34,48, 34,69, 35,27, 35,38, 35,58, 35,77 (NCH_3), 42,18, 42,27, 42,32, 42,53, 42,63 ($\text{NCH}_2\text{CHBrCH}_2\text{Br}$), 46,43, 46,62, 47,32, 47,56, 47,62, 47,95, 48,09, 48,93 ($\text{NCH}_2\text{CHBrCH}_2\text{Br}$), 53,197, 53,35, 53,77, 54,31, 54,69, 54,78, 56,31, 57,91 ($\text{NCH}_2\text{CHBrCH}_2\text{Br}$), 163,89, 164,89 ($\text{C}(\text{NR}_1\text{R}_2)_2(\text{NRR})$). **m/z (FAB⁺)** 894 (M^+ , 2,1), 893 (M^+ , 10,7), 892 (M^+ , 6,7), 891 (M^+ , 38,3), 890 (M^+ , 14,2), 889 (M^+ , 78,5), 888 (M^+ , 17,8), 887 (M^+ , 100), 885 (M^+ , 81,4), 884 (M^+ , 7,6), 883 (M^+ , 42,9), 882 (M^+ , 2,1), 881 (M^+ , 12,8) – diversas combinações de isótopos de bromo.; 807 ($\text{C}_{15}\text{H}_{26}\text{Br}_7\text{N}_3^+$, 16,8) – fragmento com combinação de isótopos maioritária, 727 ($\text{C}_{15}\text{H}_{26}\text{Br}_6\text{N}_3^+$, 34,1) - fragmento com combinação de isótopos maioritária, 645 ($\text{C}_{15}\text{H}_{26}\text{Br}_5\text{N}_3^+$, 7,6) - fragmento com combinação de isótopos maioritária. 568 ($\text{C}_{15}\text{H}_{26}\text{Br}_4\text{N}_3^+$, 7,7) - fragmento com combinação de isótopos maioritária.

II-5.4.1.3: Bis(trifluorometanosulfonil)imidato de *N,N*-dimetil-*N',N',N'',N''*-tetraalilguanidínio (12a)

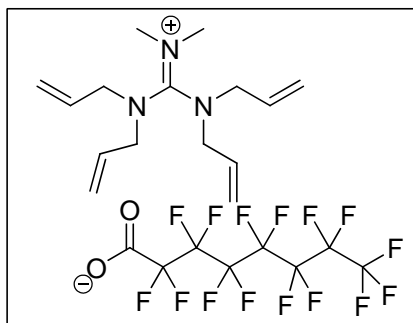


Preparado através da adição de bis(trifluorosulfonil)imidato de lítio (2,53 g, 8,82 mmol) a solução de cloreto de *N,N*-dimetil-*N',N',N'',N''*-tetraalilguanidínio (**6a**) (2 g, 7,06 mmol) em acetona (15 mL). A mistura resultante ficou em agitação contínua durante 56 h à temperatura ambiente. Na fase de *work-up* a mistura foi filtrada evaporou-se o solvente sob pressão reduzida e o resíduo foi redissolvido em diclorometano, a mistura foi novamente filtrada evaporado solvente sob pressão reduzida sendo obtido no final um líquido amarelo (3,24 g, 86,8%). *t. v.* -57,94°C, *t. d.* > 300°C ρ^a 1,342 (25°C). **IV_{vmax} (filme):** 3087 (CH_{olefínica}), 2985 (CH_{alifática}), 2925 (CH_{alifática}), 1645 (C-N), 1541 (C-N), 1527 (C-N), 1508, 1436, 1419, 1348 ((CF₃SO₂)₂N⁻), 1330, 1294, 1278, 1226, 1184 ((CF₃SO₂)₂N⁻), 1134, 1056, 993 (CH_{olefínica}), 935 (CH_{olefínica}), 877 (CH_{olefínica}), 865 (CH_{olefínica}), 788, 761, 740 cm⁻¹. **¹H-RMN (400 MHz, CDCl₃)** δ : 2,93 (6H, s, NCH₃), 3,74 (8H, sl, NCH₂CH=CH₂), 5,31-5,35 (8H, m, -CH₂CH=CH₂). 5,68 (4H, sl, -CH₂CH=CH₂) ppm **¹³C-RMN (100 MHz, CDCl₃)** δ : 40,27 (NCH₃), 52,39, 52,86 (NCH₂CH=CH₂), 114,87, 118,06 121,25, 124,44 ((CF₃SO₂)₂N⁻), 121,75 e 121,97 (NCH₂CH=CH₂, rotâmeros), 130,44 (NCH₂CH=CH₂), 162,95 (C(NR₁R₂)₂(NRR)) ppm. **Análise elementar calculada para C₁₇H₂₆N₄F₆O₄S₂·2,9H₂O (%):** N: 9,65 C: 35,16, H 5,52. **composição determinada (%):** N 9,87, C 34,67, H 5,00.

II-5.4.1.4: Bis(trifluorometanosulfonil)imidato de *N,N*-dimetil-*N',N',N'',N''*-tetraquis(2,3-dibromopropil) guanidínio (13a)


Adaptando procedimento descrito na literatura⁹² este composto foi preparado a partir de bis(trifluorometanosulfonil)imidato de *N,N*-dimetil-*N',N',N'',N''*-tetraalilguanidínio (**12a**) (1,51g, 2,86 mmol) dissolvido em diclorometano (20 mL), a essa solução foi adicionado bromo (Br₂ -3,65g, 22,87 mmol). A mistura resultante ficou em agitação contínua em refluxo durante 51 h sendo a reacção seguida por cromatografia de camada fina (*c.c.f.*), findo esse período procedeu-se ao *work-up* da reacção, nessa fase a mistura foi filtrada e a fase orgânica líquida recolhida foi lavada com 1x35 mL de solução aquosa de sulfito de sódio (20%), sendo adicionalmente lavada com 3x20 mL de água. Posteriormente o solvente da mistura reaccional foi evaporado sob pressão reduzida sendo obtido um sólido castanho (1,47 g, 44,3%). *t. v.* 10,60°C, *p.f./t.d.* 143°C, ρ^c 2,15 g/cm³. **IV_{vmax} (KBr):** 3032 (CH_{alifática}), 3003 (CH_{alifática}), 2957 (CH_{alifática}), 2930 (CH_{alifática}), 1582 (C-N), 1522, (C-N) 1458, 1431, 1344 ((CF₃SO₂)₂N⁻), 1223, 1190 ((CF₃SO₂)₂N⁻), 1132, 1045, 959, 937, 897, 793, 741, 652, 619 (f, C-Br), 569 (f, C-Br), 514 (f, C-Br) cm⁻¹. **¹H-RMN (400 MHz, CDCl₃)** δ : 3,32-3,34 (8H, m, NCH₂CBr), 3,39-3,41 (6H, m, NCH₃), 3,80-3,84 (8H, m, NCH₂CHBrCH₂Br), 3,99-4,05 ppm (4H, m, NCH₂CHBrCH₂Br). **¹³C-RMN (100 MHz, CDCl₃)** δ : 32,95, 33,08, 33,30, 33,67, 33,84 e 34,27 (NCH₃, rotâmeros), 41,94, 42,47 e 42,86 (NCH₂CHBrCH₂Br), 45,41 e 45,55 (NCH₂CHBrCH₂Br), 53,71, 53,96, 54,29, 54,60 e 57,11 (NCH₂CBr), 114,99, 118,19, 121,38 e 124,58 ((CF₃SO₂)₂N⁻), 163,85, 164,64 e 165,71 (C(NR₁R₂)₂(NRR)). **Análise elementar calculada para C₁₇H₂₆Br₈F₆N₄O₄S₂ (%):** N: 4,80 C: 17,48, H 2,24. **composição determinada (%):** N 4,40, C 17,04, H 2,49.

II-5.4.1.5: Perfluorooctanoato de *N,N*-dimetil-*N',N'',N''',N''''*-tetraalilguanidínio (14a)

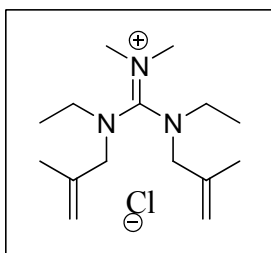


Preparado através da adição de perfluorooctanoato de sódio (1,16 g, 2,65 mmol) a solução de cloreto de *N,N*-dimetil-*N',N'',N''',N''''*-tetraalilguanidínio (**6a**) (0,5 g, 1,77 mmol) em metanol (25 mL). A mistura resultante ficou em agitação contínua à temperatura ambiente durante 56 h. Na fase de *work-up* da reacção a mistura foi filtrada sendo evaporado o solvente da fase orgânica líquida, a pressão reduzida. O resíduo foi dissolvido em diclorometano e a mistura assim obtida foi filtrada. Posteriormente o solvente da fase orgânica líquida foi evaporado a pressão reduzida, sendo obtido no final um líquido amarelo (1,02 g, 87,56%). *t. v.* -49,41, *t. d.* 160°C, ρ^b 1,37 g/cm³ (21°C). **IV_{vmax} (filme):** 3084 (CH_{olefínica}), 3015 (CH_{olefínica}), 2984 (CH_{alifática}), 2920 (CH_{alifática}), 2810 (CH_{alifática}), 1867, 1697, 1643 (C=O), 1589 (C-N), 1529 (C-N), 1420 (C=O), 1350, 1315, 1238, 1203, 1147, 1128, 1099, 1007, 993, 929, 883, 798, 744, 734, 719, 661, 640, 559, 530. **¹H-RMN (400 MHz, CDCl₃)** δ : 2,98 (6H, s, NCH₃), 3,75 (8H, sl, NCH₂CH=CH₂), 5,27-5,31 (8H, m, -CH₂CH=CH₂), 5,70 ppm (4H, sl, -CH₂CH=CH₂). **¹³C-RMN (100 MHz, CDCl₃)** δ : 40,39 (NCH₃), 52,40 e 52,80 (NCH₂CH=CH₂, rotâmeros), 110,83 (C_{fluorado}), 111,14, (C_{fluorado}) 121,45 e 122,61 (NCH₂CH=CH₂, rotâmeros), 130,87 (NCH₂CH=CH₂), 160,06, 160,29, 160,52 (-CO₂⁻), 163,06 ppm (C(NR₁R₂)₂(NRR)). **Análise elementar calculada para C₂₃H₂₆N₃O₂F₁₅ (%):** N: 6,35 C: 41,76, H 3,96. **composição determinada (%):** N 6,49, C 41,24, H 4,16.

II-5.4.2: Compostos Derivados do cloreto de *N,N*-dimetil-*N',N''*-dietil - *N',N''*-bis(2-metilalil)guanidínio (15a)

Seguindo procedimento geral descrito na secção II-5.3.1.1 - **Método geral** (para obtenção de cloretos de guanidínio) a partir da etil-metilalilamina (5g, 50,4 mmol), cloreto de fosgenimínio (3,72 g, 22,92 mmol) e 7,35 mL de trietilamina (52,7 mmol) em 50 mL de diclorometano em agitação contínua à temperatura ambiente durante 68 h. Na fase de *work-up* a mistura foi filtrada, posteriormente foi evaporado o solvente, da fase orgânica líquida recolhida, sob pressão reduzida. O resíduo resultante foi redissolvido em 35 mL de solução aquosa NaOH 2M. A partir desta fase o produto principal, cloreto de *N,N*-dimetil-*N',N''*-dietil-*N',N''*-bis(2-metilalil)guanidínio (**15a**), vai ser separado do produto secundário, 1-etil-3,3-dimetil-1-(2-metilalil)-ureia (**15b**), sendo consideradas duas secções separadas respectivamente.

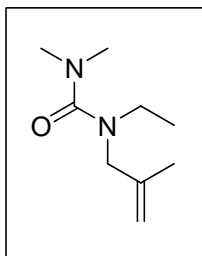
II-5.4.2.1: Cloreto de *N,N*-dimetil-*N',N''*-dietil ,*N',N''*-bis(2-metilalil)guanidínio (15a)



A partir da solução aquosa NaOH contendo resíduo da mistura reaccional, tal como aparece descrito no final da secção anterior, procedeu-se à sua lavagem com 2x40 mL de éter etílico. Posteriormente a fase aquosa foi neutralizada com solução aquosa de HCl 37%. Posteriormente foi evaporada água sob pressão reduzida. O resíduo resultante foi redissolvido em diclorometano procedendo-se a filtração. Posteriormente realizou-se evaporação de solvente sob pressão reduzida resultando um sólido amarelo, sendo adicionada acetona aparecendo precipitado branco no fundo, sendo isolado por filtração um sólido branco (2,94 g, 44,6%), *p. f.* 146,52°C, *t. d.* 213,64°C, ρ^c 1,02 g/cm³. **IV_{vmax} (KBr):** 3083 (CH_{olefinica}), 2976 (CH_{alifática}), 2939 (CH_{alifática}), 2135, 2077, 2042, 1650 (C-N), 1579 (C-N), 1537 (C-N), 1456, 1438, 1419, 1380, 1361, 1346, 1338, 1292, 1240, 1213, 1168, 1145, 1126, 1074, 1061, 1045, 1027, 989 (CH_{olefinica}), 918 (CH_{olefinica}), 862 (CH_{olefinica}), 798, 615, 580, 559, 517 cm⁻¹. **¹H-RMN (400 MHz, CDCl₃)** δ : 1,16 (6H, sl, NCH₂CH₃), 1,66 (6H, s, NCH₂C(=CH₂)CH₃), 3,07-3,18 (10H, m, NCH₃ + NCH₂CH₃), 3,81-3,91 (4H, m, NCH₂C(=CH₂)CH₃), 5,02-5,05 ppm (4H, m, NCH₂C(=CH₂)CH₃), **¹³C-RMN (100 MHz, CDCl₃)** δ : 12,96 e 13,25 (NCH₂CH₃, rotâmeros), 20,32 (NCH₂C(=CH₂)CH₃), 40,72 e 41,00 (NCH₃, rotâmeros), 44,41 e 44,72 (NCH₂CH₃, rotâmeros), 55,54 e 56,04 (NCH₂C(=CH₂)CH₃), 117,82 (NCH₂C(=CH₂)CH₃), 137,23 e

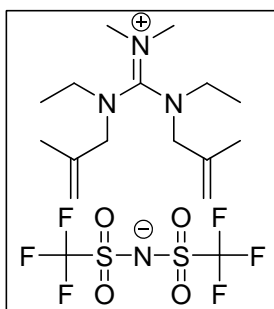
137,65 ($\text{NCH}_2\text{C}(\text{=CH}_2)\text{CH}_3$), 163,45 ppm ($\text{C}(\text{NR}_1\text{R}_2)_2(\text{NRR})$). **Análise elementar calculada para $\text{C}_{15}\text{H}_{30}\text{N}_3\text{Cl}$ (%):** N: 14,60 C: 62,58, H 10,50. **composição determinada (%):** N 14,48, C 62,01, H 10,03. **m/z (IE^+)** 252 (M^+ , 100), 99 ($\text{C}_6\text{H}_{12}\text{N}^+$, 72,51), 84 ($\text{C}_5\text{H}_{10}\text{N}^+$, 81,95) **EMAR:** 252,244117 ($\text{C}_{15}\text{H}_{30}\text{N}_3^+$ requer 252,243973).

II-5.4.2.2: 1-etil-3,3-dimetil-1-(2-metilalil)-ureia (15b)



A partir da solução aquosa NaOH contendo resíduo da mistura reaccional, tal como aparece descrito no final da secção II-5.4.2, procedeu-se à extracção com 2x40 mL de éter etílico. Depois de recolhida a fase orgânica procedeu-se à evaporação de solvente sob pressão reduzida, sendo obtido no final um líquido laranja (1,76 g, 45,16%), ρ^b 0,93 g/cm³ (23°C). **^1H -RMN (400 MHz, CDCl_3)** δ : 1,06-1,09 (3H, t, J = 7,04 Hz, NCH_2CH_3), 1,66 (3H, s, $\text{NCH}_2\text{C}(\text{=CH}_2)\text{CH}_3$), 2,78 (6H, s, NCH_3), 3,09-3,15 (2H, q, J = 7,08 Hz, NCH_2CH_3), 3,66 (2H, s, $\text{NCH}_2\text{C}(\text{=CH}_2)\text{CH}_3$), 4,87 ppm (2H, s, $\text{NCH}_2\text{C}(\text{=CH}_2)\text{CH}_3$). **^{13}C -RMN (100 MHz, CDCl_3)** δ : 12,68 (NCH_2CH_3), 20,12 ($\text{NCH}_2\text{C}(\text{=CH}_2)\text{CH}_3$), 38,59 (NCH_3), 42,34 (NCH_2CH_3), 53,47 ($\text{NCH}_2\text{C}(\text{=CH}_2)\text{CH}_3$), 111,45 ($\text{NCH}_2\text{C}(\text{=CH}_2)\text{CH}_3$), 141,89 ($\text{NCH}_2\text{C}(\text{=CH}_2)\text{CH}_3$), 165,29 ppm ($(\text{NR}_1\text{R}_2)(\text{NRR})\text{C}=\text{O}$).

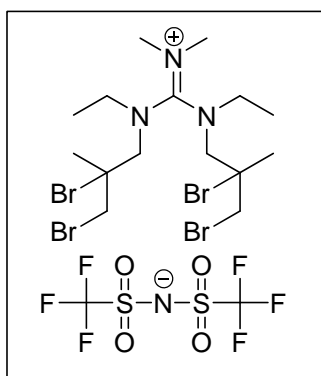
II-5.4.2.3: Bis(trifluorometanosulfonil)imidato de N,N -dimetil- N',N'' -dietil- N',N'' -bis(2-metilalil)guanidínio (16a)



Preparado a partir de cloreto de N,N -dimetil- N',N'' -dietil, N',N'' -bis(2-metilalil)guanidínio (15a) (1,5 g, 5,2 mmol) dissolvido em 20 mL de acetona. A esta solução foi adicionado bis(trifluorometanosulfonil)imidato de lítio (2,46 g, 8,9 mmol). A mistura resultante ficou em agitação contínua à temperatura ambiente durante 56 h. Na fase de *work-up* o solvente foi evaporado

sob pressão reduzida, ao resíduo resultante adicionou-se diclorometano sendo filtrada a mistura resultante. Posteriormente foi evaporado o solvente da fase líquida recolhida sendo obtido um líquido amarelo (2,45g, 70%). *t. v.* -56,4°C, *t. d.* > 300°C, ρ^b 1,28 g/cm³ (22°C). **IV_{vmax} (filme):** 3082 (CH_{olefínica}), 2978 (CH_{alifática}), 2941 (CH_{alifática}), 2814, 1845, 1655 (C-N), 1570 (C-N), 1527 (C-N), 1458, 1420, 1381, 1350 ((CF₃SO₂)₂N⁻), 1331, 1294, 1225, 1134 ((CF₃SO₂)₂N⁻), 1055, 987 (CH_{olefínica}), 914 (CH_{olefínica}), 870 (CH_{olefínica}), 820 (CH_{olefínica}), 787, 739, 652, 615, 599, 568, 513 cm⁻¹. **¹H-RMN (400 MHz, CDCl₃)** δ : 1,15-1,20 (6H, m, NCH₂CH₃), 1,70 (6H, s, NCH₂C(=CH₂)CH₃), 2,94-3,03 (6H, m, NCH₃), 3,20-3,28 (6H, m, NCH₂CH₃), 3,69-3,79 (4H, m, NCH₂C(=CH₂)CH₃), 5,07-5,11 ppm (4H, m, NCH₂C(=CH₂)CH₃). **¹³C-RMN (100 MHz, CDCl₃)** δ : 12,21 e 12,59 (NCH₂CH₃, rotâmeros), 19,15 e 19,65 (NCH₂C(=CH₂)CH₃, rotâmeros), 39,76, 40,01 e 40,22 (NCH₃, rotâmeros), 43,78 e 44,39 (NCH₂CH₃), 55,08, 55,66 (NCH₂C(=CH₂)CH₃), 114,60 ((CF₃SO₂)₂N⁻), 117,55 ((CF₃SO₂)₂N⁻), 117,80 e 117,99 (NCH₂C(=CH₂)CH₃), 120,99 ((CF₃SO₂)₂N⁻), 124,19 ((CF₃SO₂)₂N⁻), 136,72 e 136,99 (NCH₂C(=CH₂)CH₃), 162,74 e 163,15 ppm (C(NR₁R₂)₂(NRR)). **Análise elementar calculada para C₁₇H₃₀N₄O₄S₂F₆ (%):** N: 10,52 C: 38,34, H 5,68. **composição determinada (%):** N 10,51, C 38,16, H 5,56.

II-5.4.2.4: Bis(trifluorometanosulfonil)imidato de *N,N*-dimetil-*N',N''*dietil *N',N''*-bis(2,3-dibromo-2-metil-propil) guanidínio (17a)



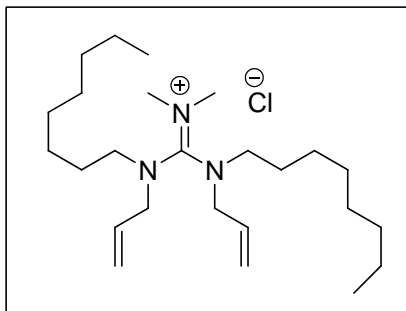
Adaptando procedimento descrito na literatura⁹² este composto foi preparado a partir de bis(trifluorometanosulfonil)imidato de *N,N*-dimetil-*N',N''*dietil *N',N''*-bis(2-metilalil)guanidínio (**16a**) (1,76 g, 3,31 mmol) dissolvido em diclorometano (25 mL), a essa solução foi adicionado bromo (2,12 g, 13,2 mmol), a mistura assim obtida ficou a reagir durante 56 horas em refluxo. Findo esse período procedeu-se ao *work-up* da reacção, nessa fase a mistura foi filtrada e a fase líquida recolhida foi lavada com 1x25 mL de solução aquosa de sulfito de sódio (20%) e posteriormente lavada com 2x25 mL de água. À fase orgânica posteriormente adicionou-se sulfato de magnésio sendo seguidamente filtrada. O solvente da fase líquida recolhida foi evaporado, resultando no final sólido amarelo (2,00 g, 75%), *t. v.* -10,41°C, *p. f.* 74,31°C, *t. d.* 161°C, ρ^c 1,68 g/cm³. **IV_{vmax} (filme):** 3038 (CH_{alifática}), 2995 (CH_{alifática}), 2977 (CH_{alifática}), 2945 (CH_{alifática}), 2588, 2468, 2365, 2316, 2270, 2190,

2154, 2119, 2011, 1961, 1928, 1880, 1589 (C-N), 1577 (C-N), 1506, 1452, 1419, 1386, 1348 ((CF₃SO₂)₂N⁻), 1296, 1259, 1224, 1174 ((CF₃SO₂)₂N⁻), 1134, 1054, 1045, 1003, 980, 933, 897, 869, 834, 764, 727, 715, 651, 635 (f, C-Br) cm⁻¹. **¹H-RMN (400 MHz, CDCl₃)** δ: 1,33-1,43 (6H, m, NCH₂CH₃), 1,87-2,01 (6H, m, NCH₂C(CH₃)(CBrCH₂Br)), 2,97-2,99 (4H, m, NCH₂CH₃), 3,27-3,29 (6H, m, NCH₃), 3,66-3,71 (4H, m, NCH₂C(CH₃)(CBrCH₂Br)), 3,79-3,92 ppm (4H, m, NCH₂C(CH₃)(CBrCH₂Br)). **¹³C-RMN (100 MHz, CDCl₃)** δ: 12,34-14,05 (NCH₂CH₃), 29,08, 29,26, 29,31 (NCH₂C(CH₃)(CBrCH₂Br)), 39,92 (NCH₃), 40,89, 41,60, 41,82 (NCH₂CH₃), 47,96, 48,11, 48,22, 48,59 (NCH₂C(CH₃)(CBrCH₂Br)), 54,20, 54,79, 54,97, 57,02, 57,24 (NCH₂C(CH₃)(CBrCH₂Br)), 65,01, 66,79, 66,99, 67,27 (NCH₂C(CH₃)(CBrCH₂Br)), 115,06, 118,25, 121,45, 124,65 ((CF₃SO₂)₂N⁻), 166,81, 167,42 ppm (C(NR₁R₂)₂(NRR)). **Análise elementar calculada para C₁₇H₃₀N₄O₄S₂F₆Br₄ (%):** N: 6,57 C: 23,96, H 3,55. **composição determinada (%):** N 6,57, C 23,93, H 3,85. **m/z (ESI-TOF)** 577 (M⁺, 1,5), 576 (M⁺, 20,8), 575 (M⁺, 11,7), 574 (M⁺, 83,2), 573 (M⁺, 19,7), 572 (M⁺, 100), 571 (M⁺, 14,3), 570 (M⁺, 82,2), 569 (M⁺, 2,1), 568 (M⁺, 24,0). **EMAR:** 571,9135 (C₁₅H₃₀Br₄N₃⁺ requer 571,9132).

II-5.4.3: Síntese do cloreto de *N,N* - dimetil – *N',N''*-dialil-*N',N''*-dioctilguanidínio (18a)

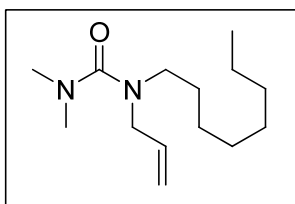
Seguindo procedimento geral descrito na secção II-5.3.1.1 - **Método geral** (para obtenção de cloretos de guanidínio) a partir da e aliloctilamina (1g, 5,9 mmol) (preparada previamente em II-5.4.3.3), cloreto de fosgeniminio (0,46 g, 2,81 mmol) e 0,9 mL de trietilamina em 7,5 mL de diclorometano em agitação contínua à temperatura ambiente durante 19 h. Na fase de *work-up* a mistura reaccional foi filtrada e a fase líquida recolhida foi submetida a pressão reduzida de forma a eliminar solvente, ao resíduo resultante foi adicionado água formando-se duas fases que foram decantadas, a fase orgânica (fase superior) foi submetida a pressão reduzida sendo eliminado o solvente. Posteriormente redissolvido o resíduo no mínimo de diclorometano, este foi adicionado a uma coluna cromatográfica (c.c.) com sílica gel como fase estacionária, foram obtidas duas fracções principais.

II-5.4.3.1: Cloreto de *N,N* - dimetil – *N',N''*-dialil-*N',N''*diocetilguanidínio (18a)



No seguimento do procedimento descrito na secção anterior depois de realizar uma cromatografia de coluna, a segunda fracção a ser eluída depois de eliminar solvente sob pressão reduzida resultou num líquido viscoso laranja (0,35 g, 14%), *t. v.* -53,26°C, *t. d.* 195,42°C. **¹H-RMN (400 MHz, CDCl₃)** δ: 0,58-0,67 (6H, m, -CH₂CH₃), 0,97-1,05 (24H, m, -CH₂-), 2,90 (6H, s, NCH₃), 3,01-3,03 (4H, m, NCH₂C₇H₁₅), 3,53-3,69 (4H, m, NCH₂CH=CH₂), 5,13-5,24 (4H, m, NCH₂CH=CH₂), 5,49-5,67 (2H, m, NCH₂CH=CH₂). **¹³C-RMN (100 MHz, CDCl₃)** δ: 13,52 (-CH₂CH₃), 22,04 (-CH₂-), 26,25, 26,46 e 26,75 (-CH₂-), 27,20 (-CH₂-), 28,58 (-CH₂-), 30,63 e 31,17 (-CH₂-), 40,24 e 40,61 (NCH₃), 48,84, 49,27, 49,56 e 49,74 (NCH₂CH₂-), 52,18, 52,43, 52,70, 53,17 (NCH₂CH=CH₂), 120,80, 121,06 (NCH₂CH=CH₂), 130,39, 130,93 (NCH₂CH=CH₂), 162,85 ppm (C(NR₁R₂)₂(NRR)). **m/z (IE⁺)** 392 (M⁺, 77,79), 238 (C₁₄H₂₈N₃⁺, 37,53), 224 (C₁₄H₂₈N₂⁺, 58,12), 183 (C₁₁H₂₃N₂⁺, 57,08), 71 (C₄H₉N⁺, 100) **EMAR:** 392,4005 (C₂₅H₅₀N₃⁺ requer 392,3999).

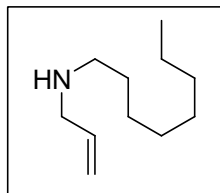
II-5.4.3.2: 1-alil-3,3-dimetil-1-octilureia (18b)



No seguimento do procedimento descrito na secção **II-5.4.3** depois de realizar uma cromatografia de coluna (*c.c.*), a primeira fracção a ser eluída depois de eliminar solvente sob pressão reduzida resultou num líquido laranja (0,074g, 11%). **¹H-RMN (400 MHz, CDCl₃)** δ: 0,82-0,85 (3H, t, *J* = 7,00 Hz, -CH₂CH₃), 1,22 (12H, sl, -CH₂-), 2,76 (6H, s, NCH₃), 3,02-3,06 (2H, t, *J* = 7,68 Hz, NCH₂CH₂-), 3,70-3,71 (2H, d, *J* = 5,68 Hz, NCH₂CH=CH₂), 5,09-5,17 (2H, m, NCH₂CH=CH₂), 5,72-5,81 ppm (1H, m, NCH₂CH=CH₂). **¹³C-RMN (100 MHz, CDCl₃)** δ: 13,99 (-CH₂CH₃), 22,55 (-CH₂CH₃), 26,93 (-CH₂-), 27,51 (-CH₂-), 29,17 (-CH₂-), 31,60 (-CH₂-), 38,63 (NCH₃), 47,70 (NCH₂CH₂-), 50,98 (NCH₂CH=CH₂), 116,61 (NCH₂CH=CH₂), 134,81 (NCH₂CH=CH₂), 165,22 ppm

((NR₁R₂)(NRR)C=O). **Análise elementar calculada para C₁₄H₂₈N₂O (%):** N: 11,65 C: 69,95, H 11,74. **composição determinada (%):** N 12,56, C 70,23, H 11,34.

II-5.4.3.3: Alil-octilamina (18c)



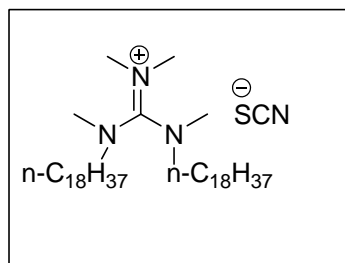
Preparado a partir de alilamina (12,28 g, 215,2 mmol) ao qual se adicionou carbonato de sódio e clorooctano (4 g, 26,9 mmol) a mistura esteve em agitação a 80°C num micro-reactor durante 36 h. Na fase de *work-up* filtrou-se a mistura e a solução resultante foi sujeita a pressão reduzida para evaporar alilamina que havia sido colocada em grande excesso. No final foi obtido um líquido translúcido (1,99 g, 43,7%). **¹H-RMN (400 MHz, CDCl₃)** δ: 0,86-0,89 (3H, t, J = 6,8 Hz, -CH₂CH₃), 1,27-1,29 (10H, m, -CH₂-), 1,48-1,53 (2H, m, NCH₂CH₂-), 2,29 (1H, sl, NH(CH₂)₂), 2,60-2,64 (2H, t, J = 7,2 Hz, NCH₂CH₂-), 3,27-3,28 (2H, d, J = 6 Hz, NCH₂CH=CH₂), 5,10-5,12 (1H, d, J = 10,4 Hz, CHH_{cis}=CHCH₂N), 5,17-5,21 (1H, d, J = 17,2 Hz, CHH_{trans}=CHCH₂N), 5,88-5,98 (1H, m, CHH=CHCH₂N). **¹³C-RMN (100 MHz, CDCl₃)** δ: 14,03 (-CH₂CH₃), 22,60 (-CH₂CH₃), 27,31 (-CH₂-), 29,21, 29,46 e 29,79 (-CH₂-), 31,78 (-CH₂-), 49,23 (NCH₂CH₂-), 52,28 (NCH₂CH=CH₂), 116,20 (NCH₂CH=CH₂), 136,34 ppm (NCH₂CH=CH₂).

II-5.5: Síntese de Sais de Guanidínios: Baseados no Anião Dicianamida, Tiocianato e para-Toluenosulfonato

Partindo dos cloretos dos respectivos guanidínios, preparados de acordo com método geral (II-5.3.1.1), foi realizada troca de anião a partir de diversos sais, explicitado o método utilizado para cada caso particular.

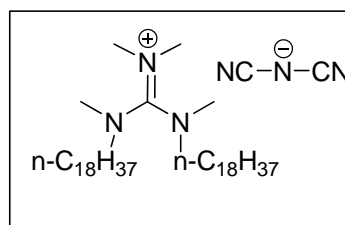
II-5.5.1: Síntese de Sais Baseados no Catião de *N, N, N',N''*-Tetrametil-*N',N'*-dioctadecilguanidínio

II-5.5.1.1: Tiocianato de *N, N, N',N''*-Tetrametil-*N',N'*-dioctadecilguanidínio (19a)



Numa mistura de 60 mL em acetato de etilo/acetona (1:1 volume) estando dissolvido cloreto de *N, N, N',N''*-Tetrametil-*N',N'*-dioctadecilguanidínio (**4a**) (0,5 g, 0,722 mmol) foi adicionado tiocianato de potássio (0,14 g, 1,44 mmol). A mistura ficou em agitação à temperatura ambiente durante 48 h. Na fase de *work-up* os solventes foram evaporados a pressão reduzida e o resíduo redissolvido em diclorometano. A mistura foi filtrada e a fase líquida filtrada foi submetida a pressão reduzida de forma a eliminar solvente resultando um sólido branco (0,5 g, 83,6%). *p.f.* 83,49°C. **IV_{vmax} (KBr):** 2920 (CH_{alifática}), 2850 (CH_{alifática}), 2060 (S=C=N), 1585 (C-N) e 1562 (C-N), 1466, 1412, 1369, 1254, 1180, 1157, 1134, 1088, 1057, 918, 895, 852, 721 cm⁻¹. **¹H-RMN (400 MHz, CDCl₃)** δ: 0,83-0,866 (6H, t, J = 6,6, CCH₃), 1,23 (64H, s, CCH₂C) 2,99-3,11(12H, m, NCH₃), 3,23 (4H, sl, NCH₂C). **¹³C-RMN (100 MHz, CDCl₃)** δ: 14,02 (CCH₃), 22,59 (CCH₂C), 26,74 (CCH₂C), 27,62 (CCH₂C), 29,20, 29,27 (CCH₂C), 29,62, 31,84 (CCH₂C), 38,12 (NCH₃), 40,36, 40,78(NCH₃, rotâmeros), 52,87 (NCH₂C), 163,68 (C(NR₁R₂)₂(NRR)). **Análise elementar calculada para C₄₂H₈₆N₄S (%):** N 8,25, C 74,27, H 12,76; **composição determinada (%):** N 8,39, C 74,30, H 12,92.

II-5.5.1.2: Dicianamida de *N, N, N',N''*-Tetrametil-*N',N'*-dioctadecilguanidínio (20a)

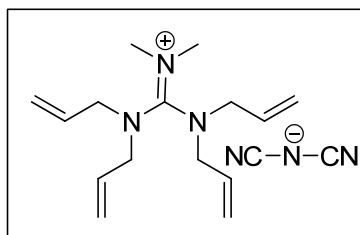


Estando dissolvido cloreto de *N, N, N',N''*-Tetrametil-*N',N'*-dioctadecilguanidínio (**4a**) (0,15g, 0,22 mmol) em 15 mL de diclorometano foi adicionado dicianamida sódica (0,029g, 0,325 mmol). A mistura ficou em agitação durante 24h à temperatura ambiente. Na fase de *work-up* a mistura foi filtrada e a fase orgânica líquida foi submetida a pressão reduzida para remoção de

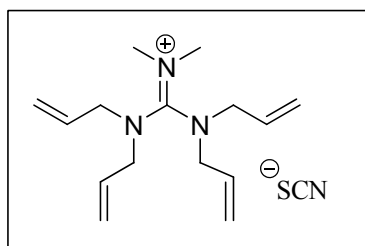
solvente sendo obtido um sólido branco (0,10g, 67,1%). *p.f.* 77,97 °C. **IV_{vmax} (KBr):** 3132 (fr), 2920 (f, CH_{alifática}), 2850 (f, CH_{alifática}), 2353, 2225 (f, NC-N-CN), 2187 (f, NC-N-CN), 2133 (f, NC-N-CN), 1570 (f, C-N), 1466, 1412, 1365, 1304, 1254, 1184, 1138, 1088, 1061, 949, 918, 895, 852, 721, 663, 521 cm⁻¹. **¹H-RMN (400 MHz, CDCl₃)** δ: 0,86-0,90 (6H, t, J = 6,4 Hz, CCH₃), 1,255 (64H, s, CCH₂C) 2,97-3,09 (12H, m, NCH₃), 3,21 (4H, sl, NCH₂C). **¹³C-RMN (100 MHz, CDCl₃)** δ: 14,06 (CCH₃), 22,63 (CCH₂CH₃), 26,75 (CCH₂C), 27,64 (CCH₂C), 29,21, 29,31, 29,53, 29,66 (CCH₂C, rotâmeros), 31,87 (CCH₂C), 38,11 e 38,38 (NCH₃, rotâmeros), 40,36, 40,73 (NCH₃, rotâmeros), 52,93 (NCH₂C), 119,93 (NCNCN), 163,65 (C(NR₁R₂)₂(NRR)). **Análise elementar calculada para C₄₃H₈₆N₆ (%):** N 12,23, C 75,16, H 12,61; **composição determinada (%):** N 12,23, C 74,99, H 12,83.

II-5.5.2: Síntese de Sais Baseados no Catião *N,N*-dimetil-*N',N'',N''',N''''*-tetraalilguanidínio

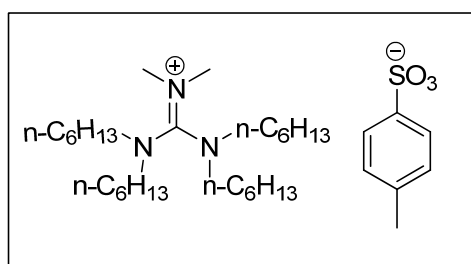
II-5.5.2.1: Dicianamida de *N,N*-dimetil-*N',N'',N''',N''''*-tetraalilguanidínio (21a)



A uma solução de cloreto de *N,N*-dimetil-*N',N'',N''',N''''*-tetraalilguanidínio (**6a**) (0,5 g, 2,12 mmol) em 20 mL de diclorometano foi adicionada dicianamida sódica (0,236 g, 2,65 mmol). A mistura ficou em agitação durante 72 h à temperatura ambiente. Na fase de *work-up* a mistura foi filtrada, e, depois de recolhidas as fases líquidas combinadas, o solvente foi removido sobre pressão reduzida. No final obteve-se um sólido amarelo (0,487 g, 73,13%). *p. f.* 50°C., *t. d.* 225°. **IV_{vmax} (KBr):** 3086 (CH_{olefínica}), 3000 (CH_{alifática}), 2926 (CH_{alifática}), 2293, 2229, 2195 e 2137 (NC-N-CN), 1644 (C-N), 1585 (C-N), 1538 (C-N), 1417, 1347, 1315, 1234, 1144, 1062, 996 (CH_{olefínica}), 931 (CH_{olefínica}), 880 cm⁻¹ (CH_{olefínica}). **¹H-RMN (400 MHz, CDCl₃)** δ: 3,00 (6H, s, NCH₃), 3,77-3,85 (8H, m, NCH₂CH=CH₂), 5,32-5,36 (8H, m, -CH₂CH=CH₂), 5,74 (4H, sl, -CH₂CH=CH₂). **¹³C-RMN (100 MHz, CDCl₃)** δ: 40,48 (NCH₃), 52,44 e 52,87 (NCH₂CH=CH₂, rotâmeros), 119,82 (NC-N-CN), 121,89 e 122,08 (-CH₂CH=CH₂, rotâmeros), 130,57 (-CH₂CH=CH₂), 162,97 (C(NR₁R₂)₂(NRR)). **Análise elementar calculada para C₁₇H₂₆N₆ (%):** N 26,73, C 64,94, H 8,33; **composição determinada (%):** N 26,75, C 64,64, H 8,86.

II-5.5.2.2: Tiocianato de *N,N*-dimetil-*N',N',N'',N''*-tetraalilguanidínio (22a)


A uma solução de cloreto de *N,N*-dimetil-*N',N',N'',N''*-tetraalilguanidínio (**6a**) (0,5 g, 1,77 mmol) em 20 mL de diclorometano foi adicionado tiocianato de potássio (0,343 g, 3,5 mmol). A mistura resultante ficou em agitação durante 72 h. Posteriormente procedeu-se à fase de *work-up*, filtrando a mistura e evaporando o solvente sob pressão reduzida, resultando no final um sólido laranja (0,318 g, 58,82%). *p. f.* 71,5°C, *t. d.* 250°C. **IV_{vmax} (KBr):** 3086 (CH_{olefinica}), 3000 (CH_{alifática}), 2914 (CH_{alifática}), 2059 (SCN), 1664 (C-N), 1585 (C-N), 1539 (C-N), 1416, 1271, 1235, 1158, 1121, 1062, 993 (CH_{olefinica}), 933 (CH_{olefinica}), 882 (CH_{olefinica}), 868 (CH_{olefinica}), 727, 671, 642 cm⁻¹. **¹H-RMN (400 MHz, CDCl₃)** δ: 3,07 (6H, s, NCH₃), 3,79-3,96 (8H, m, NCH₂CH=CH₂), 5,31-5,34 (8H, m, -CH₂CH=CH₂), 5,79 (4H, sl, -CH₂CH=CH₂). **¹³C-RMN (100 MHz, CDCl₃)** δ: 40,65 (NCH₃), 52,49 e 52,83 (NCH₂CH=CH₂, rotâmeros), 121,68 (-CH₂CH=CH₂), 130,98 (-CH₂CH=CH₂), 163,01 (C(NR₁R₂)₂(NRR)). **Análise elementar calculada para C₁₆H₂₆N₄S.O,1H₂O (%):** N 18,17, C 62,34, H 8,57; **composição determinada (%):** N 18,60 C 62,29, H 8,79.

II-5.5.3: *para*-Toluenosulfonato de *N,N*-dimetil-*N',N',N'',N''*-tetrahexilguanidínio (23a)


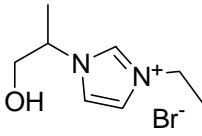
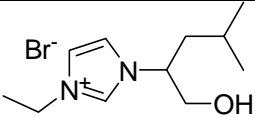
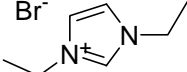
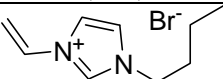
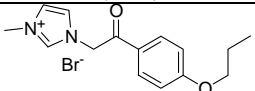
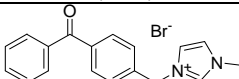
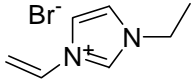

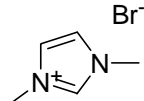
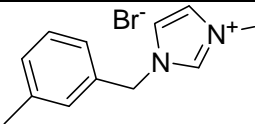
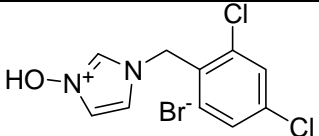
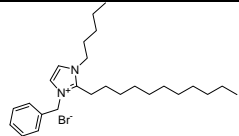
A uma solução de cloreto de *N,N*-dimetil-*N',N',N'',N''*-tetrahexilguanidínio (**8b**) (1 g, 2,15 mmol) em diclorometano (20 mL) foi adicionado *para*-toluenosulfonato de sódio (0,626 g, 3,22 mmol). A mistura ficou em agitação à temperatura ambiente durante 24 h. Findo esse tempo procedeu-se ao *work-up*, no qual a mistura foi filtrada e a correspondente fase líquida foi evaporada a pressão reduzida de forma a eliminar solvente. No final foi obtido um líquido viscoso amarelo (1,30 g, 98%). *t. v.* -66,55°C. **IV_{vmax} (filme):** 2927 (CH_{alifática}), 2858 (CH_{alifática}), 1643 (C-N), 1581 (C-N), 1539 (C-N), 1462, 1377, 1315, 1219 (RSO₃⁻), 1122, 1061 e 1034 (RSO₃⁻), 1011, 968, 906, 879, 818

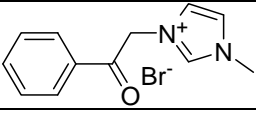
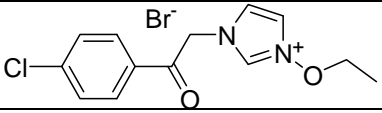
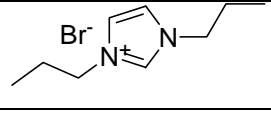
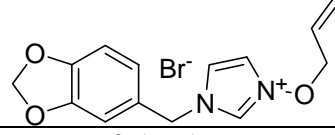
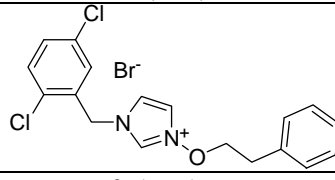
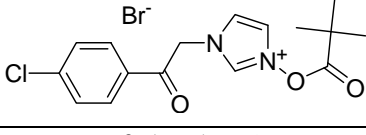
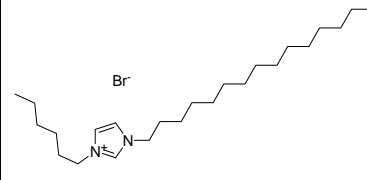
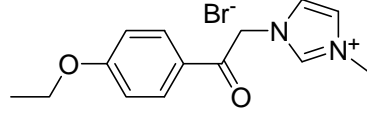
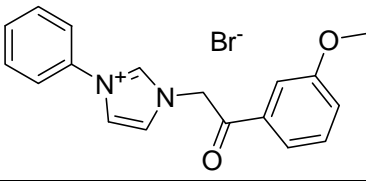
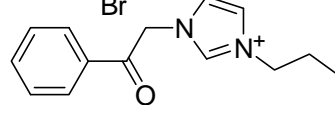
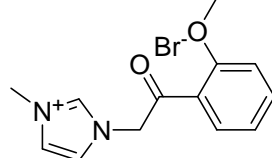
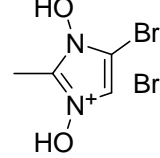
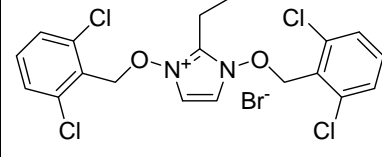
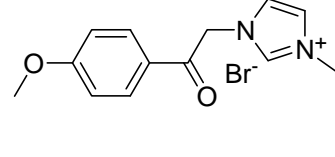
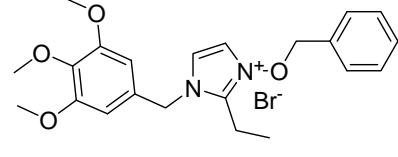
(CH_{aromática}), 729 (CH_{aromática}), 683 (CH_{aromática}), 567 cm⁻¹. **¹H-RMN (400 MHz, CDCl₃)** δ: 0,78-0,85 (12H, m, CCH₃) 1,22-1,66 (32H, m, CCH₂C), 2,29 (C_{aromático}CH₃), 3,03 (6H, s, NCH₃), 3,08-3,17 (8H, m, NCH₂C), 7,07-7,09 (2H, d, J = 7,7 Hz, CCH₃CH_{aromático}), 7,73-7,74 (2H, d, 7,8 Hz, CH_{aromático}CSO₃). **¹³C-RMN (100 MHz, CDCl₃)** δ: 13,84 (CCH₃), 21,15 (C_{aromático}CH₃), 22,39 (CCH₂CH₃), 26,34 e 26,47 (CCH₂C, rotâmeros), 27,36, 27,69 (CCH₂C, rotâmeros), 31,22 e 31,39 (CCH₂C, rotâmeros), 40,71 (NCH₃), 49,27 e 49,80 (NCH₂C, rotâmeros), 125,98 (C_{aromático}SO₃⁻), 128,30 (C_{aromático}), 138,83 (C_{aromático}), 143,72 (C_{aromático}), 163,34 (C(NR₁R₂)₂(NRR)). **Análise elementar calculada para C₃₄H₆₅N₃O₃S (%):** N 7,05, C 68,52, H 10,99; **composição determinada (%):** N 6,87 C 68,26 H 11,46.

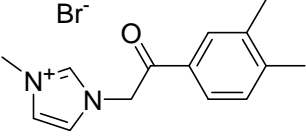
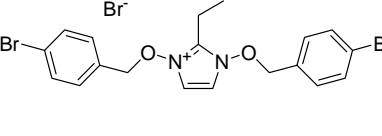
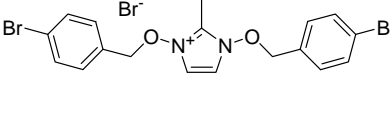
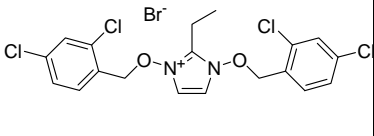
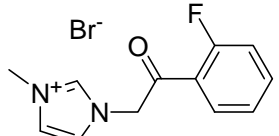
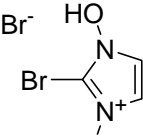
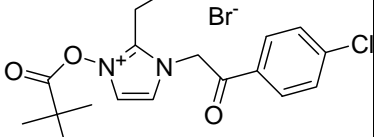
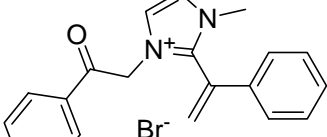
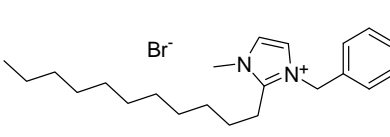
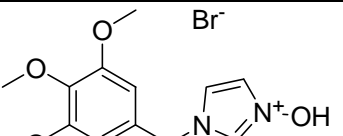
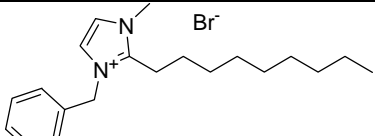
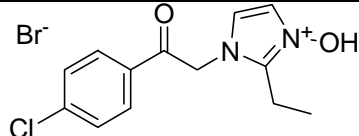
II-6 – Anexos:

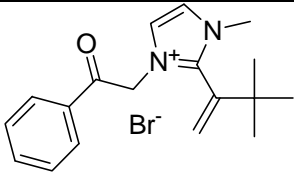
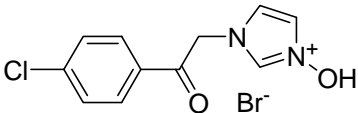
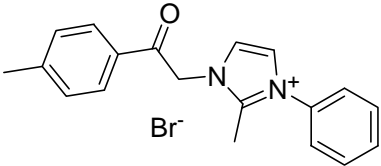
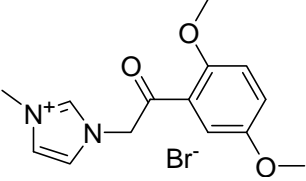
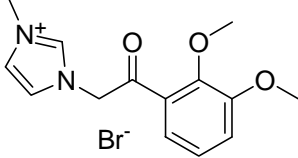
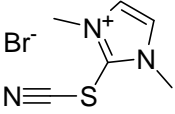
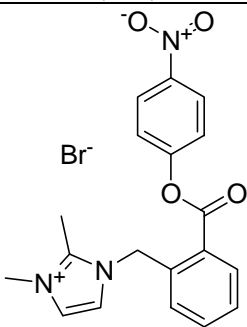
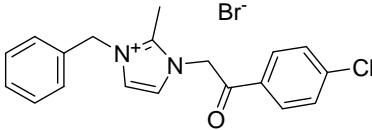
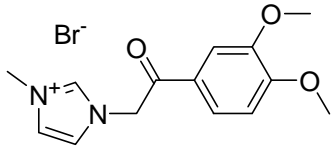
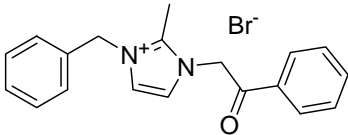
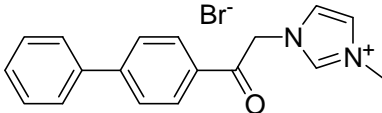
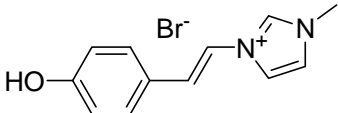
Estudo b) : Classificação de sais de Imidazólio segundo ponto de fusão/estado físico (modelo SVM)

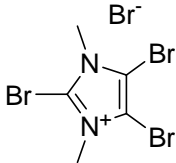
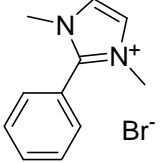
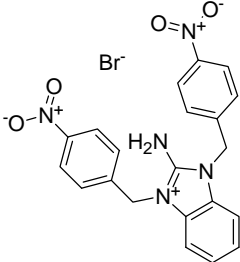
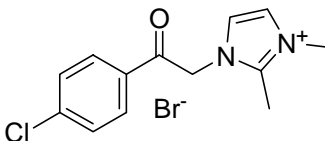
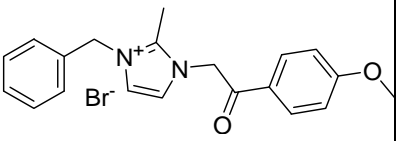
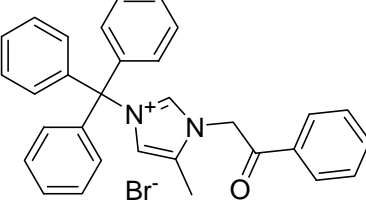
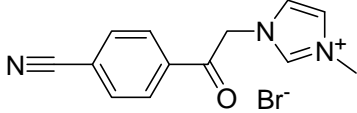
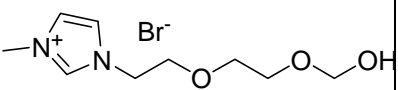
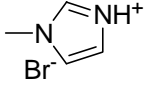
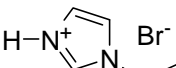
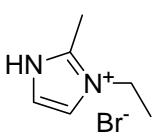

Conjunto de treino

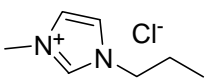
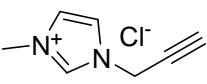
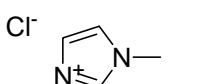
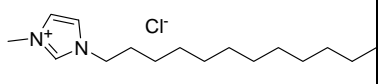
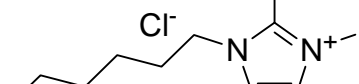
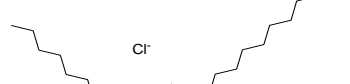
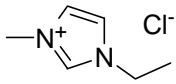
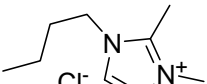
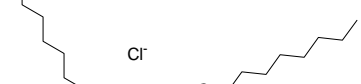
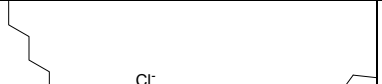


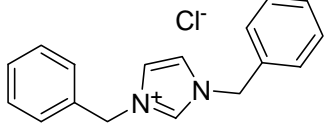
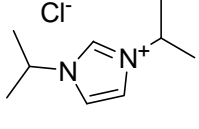
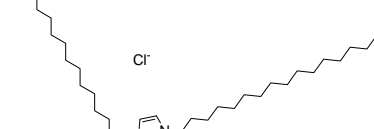
		
p.f. (exp): 5,5	p.f. (exp): 15,5	p.f. (exp): 53
a (exp)	a (exp)	b (exp)
a (prev)	a (prev)	b (prev)
<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225
		
p.f. (exp): 83	p.f. (exp): 85	p.f. (exp): 93,5
b (exp)	b (exp)	b (exp)
b (prev)	b (prev)	b (prev)
<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225
		
p.f. (exp): 99,5	p.f. (exp): 103	p.f. (exp): 110
b (exp)	c (exp)	c (exp)
b (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225
		
p.f. (exp): 115	p.f. (exp): 115	p.f. (exp): 116,5
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225

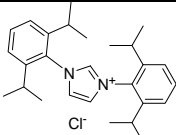
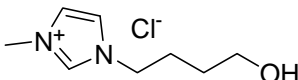
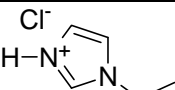
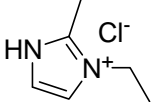
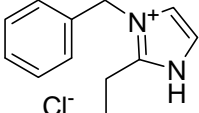
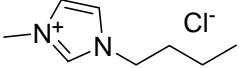
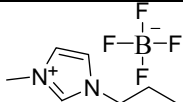
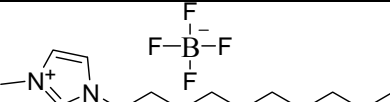
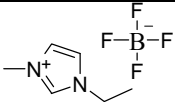
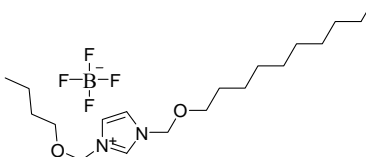
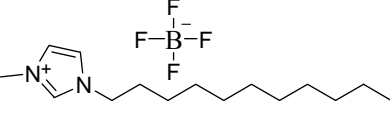
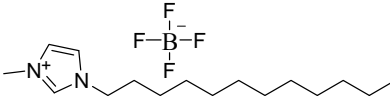
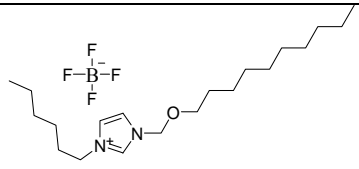
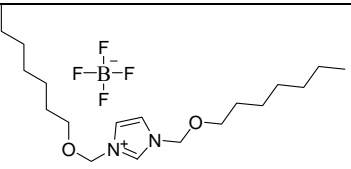
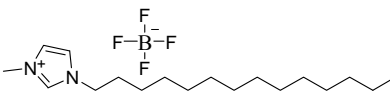
		
p.f. (exp): 117	p.f. (exp): 130	p.f. (exp): 135
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225
		
p.f. (exp): 137	p.f. (exp): 138	p.f. (exp): 140
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225
		
p.f. (exp): 142,6	p.f. (exp): 144,5	p.f. (exp): 147
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225
		
p.f. (exp): 149	p.f. (exp): 166	p.f. (exp): 154
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225
		
p.f. (exp): 156	p.f. (exp): 157,5	p.f. (exp): 159
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225

		
p.f. (exp): 164	p.f. (exp): 170	p.f. (exp): 174
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225
		
p.f. (exp): 176	p.f. (exp): 179	p.f. (exp): 181
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225
		
p.f. (exp): 184	p.f. (exp): 185	p.f. (exp): 186,5
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225
		
p.f. (exp): 190	p.f. (exp): 194	p.f. (exp): 199
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225

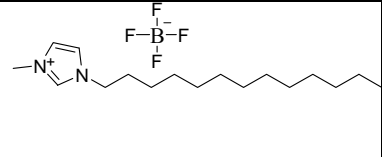
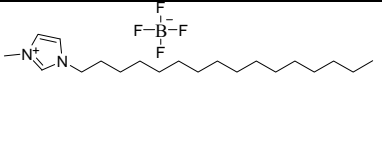
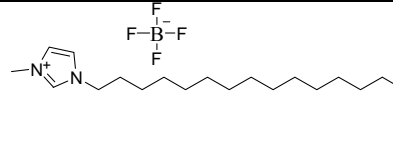
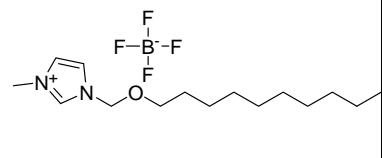
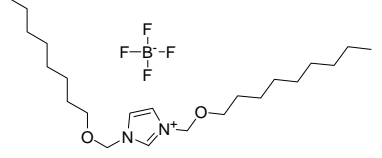
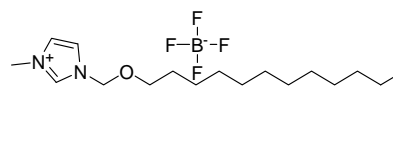
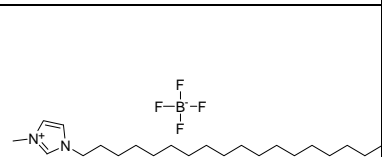
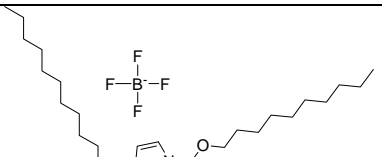
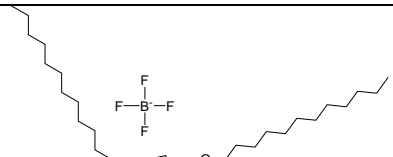
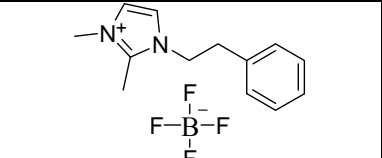
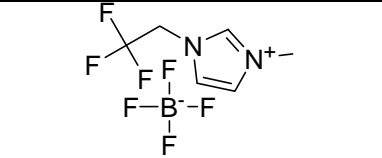
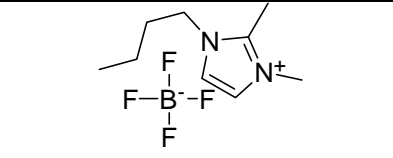
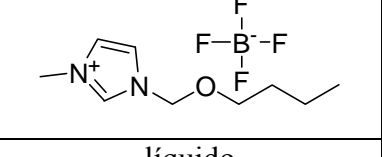
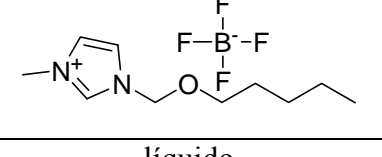
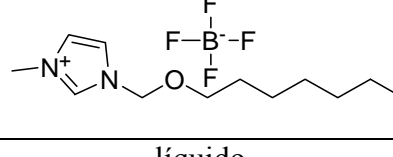
		
p.f. (exp): 200	p.f. (exp): 202	p.f. (exp): 205,5
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225
		
p.f. (exp): 208,5	p.f. (exp): 216	p.f. (exp): 220,5
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225
		
p.f. (exp): 223,5	p.f. (exp): 228,5	p.f. (exp): 231,5
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225
		
p.f. (exp): 239	p.f. (exp): 248,5	p.f. (exp): 260,5
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225

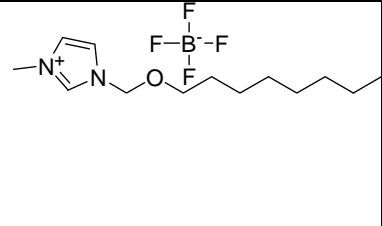
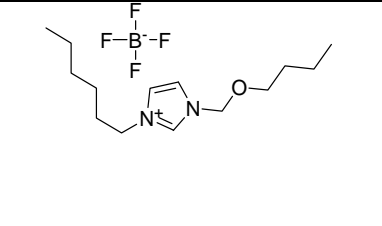
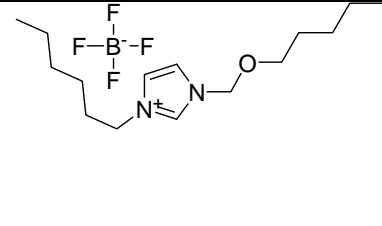
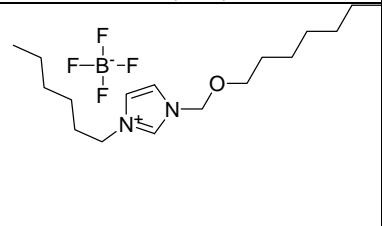
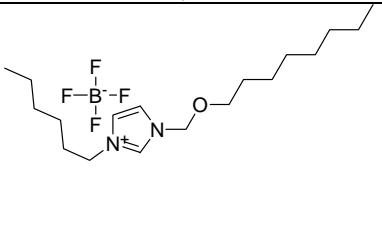
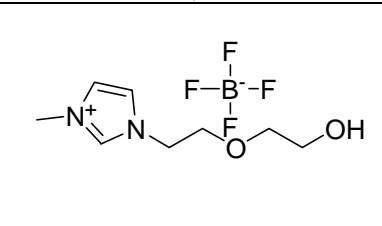
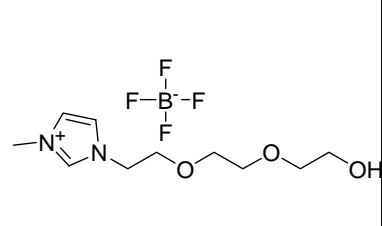
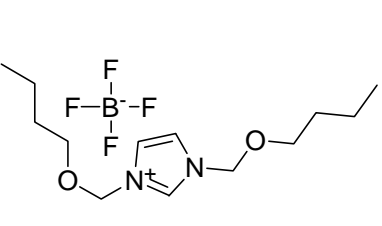
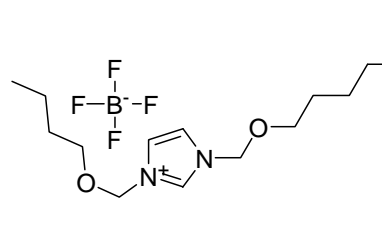
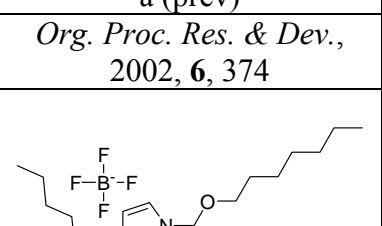
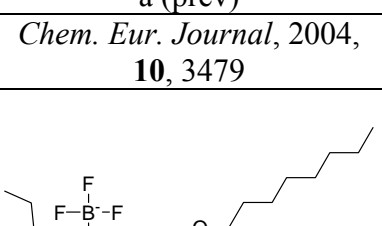
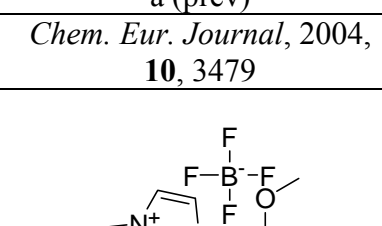
		
p.f. (exp): 268	p.f. (exp): 283	p.f. (exp): 370
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225
		
p.f. (exp): 201	p.f. (exp): 202	p.f. (exp): 205,5
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225
		
p.f. (exp): 212	líquido	p.f. (exp): 41
c (exp)	a (exp)	b (exp)
c (prev)	a (prev)	b (prev)
<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> ; 2002, 42 , 225	<i>Sol. State Ionics</i> ; 2002, 154 , 303
		
p.f. (exp): 60	p.f. (exp): 22	p.f. (exp): 123
b (exp)	c (exp)	c (exp)
b (prev)	c (prev)	c (prev)
<i>Sol. State Ionics</i> ; 2002, 154 , 303	<i>Sol. State Ionics</i> ; 2002, 154 , 303	<i>Sol. State Ionics</i> ; 2002, 154 , 303

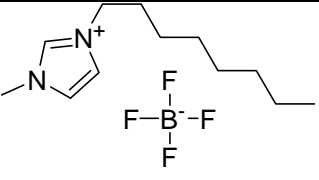
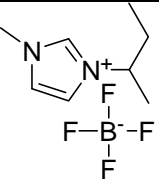
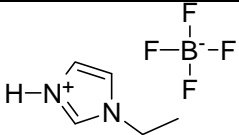
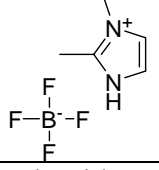
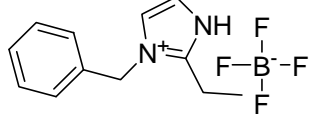
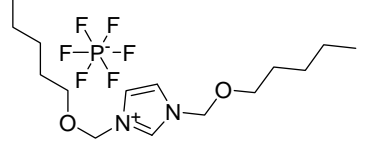
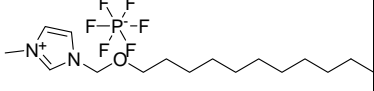
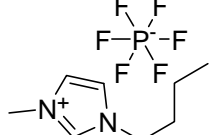
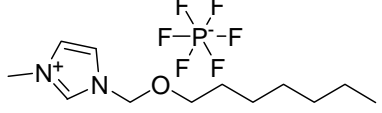
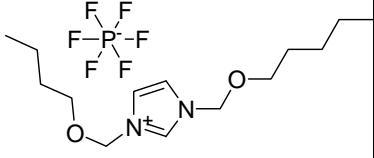
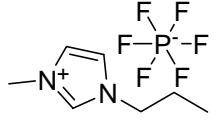
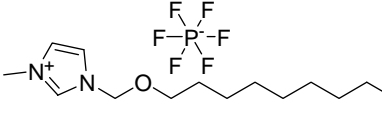
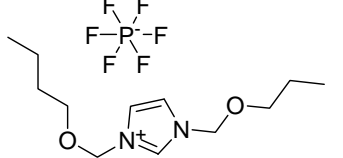
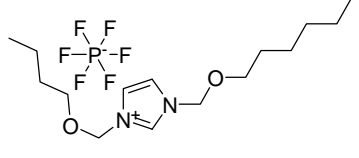
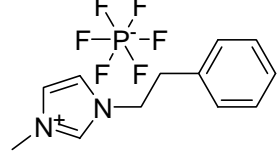
		
p.f. (exp): 62	p.f. (exp): 62	p.f. (exp): 72
b (exp)	b (exp)	b (exp)
b (prev)	b (prev)	b (prev)
Merck	<i>Chem. Comm.</i> , 1985, 845	<i>Sol. State Ionics</i> ; 2002, 154 , 303
		
p.f. (exp): 38	p.f. (exp): 44	p.f. (exp): 82
b (exp)	b (exp)	b (exp)
b (prev)	b (prev)	b (prev)
Merck	Merck	Merck
		
p.f. (exp): 83	p.f. (exp): 91	p.f. (exp): 94,5
b (exp)	b (exp)	b (exp)
b (prev)	b (prev)	b (prev)
<i>Chem. Comm.</i> , 2001, 2399	Merck	<i>Chem. Eur. Journal</i> ; 2004, 10 , 3479
		
p.f. (exp): 96	p.f. (exp): 105	p.f. (exp): 109,25
b (exp)	c (exp)	c (exp)
b (prev)	c (prev)	c (prev)
<i>Chem. Eur. Journal</i> ; 2004, 10 , 3479	<i>Chem. Eur. Journal</i> ; 2004, 10 , 3479	<i>Chem. Eur. Journal</i> ; 2004, 10 , 3479
		
p.f. (exp): 120,5	p.f. (exp): 135	p.f. (exp): 151,5
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>Synthesis</i> , 1996, 697	<i>Synthesis</i> , 1996, 697	Beilstein

		
p.f. (exp): 278	líquido	p.f. (exp): 58
c (exp)	a (exp)	b (exp)
c (prev)	a (prev)	b (prev)
Iolitec	<i>Chem. Eur. J.</i> , 2002, 8 , 3671	<i>Sol. State Ionics</i> ; 2002, 154 , 303
		
p.f. (exp): 178	p.f. (exp): 171	p.f. (exp): 62
c (exp)	c (exp)	b (exp)
c (prev)	c (prev)	b (prev)
<i>Sol. State Ionics</i> ; 2002, 154 , 303	<i>Sol. State Ionics</i> ; 2002, 154 , 303	<i>Green Chem.</i> , 2001, 3 , 156
		
p.f. (exp): -5	p.f. (exp): -4,2	p.f. (exp): 12,25
a (exp)	a (exp)	a (exp)
a (prev)	a (prev)	a (prev)
<i>Dalton.Trans.</i> , 1999, 2133	Beilstein	<i>Chem. Eur. J.</i> ; 2004, 10 , 6581
		
p.f. (exp): 16	p.f. (exp): 21,4	p.f. (exp): 26,4
a (exp)	a (exp)	b (exp)
a (prev)	a (prev)	b (prev)
<i>Ind. Chem. Res.</i> , 2001, 11 , 40	Beilstein	Beilstein
		
p.f. (exp): 38	p.f. (exp): 39,5	p.f. (exp): 43,5
b (exp)	b (exp)	b (exp)
b (prev)	a (prev)	b (prev)
<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Chem. Eur. Journal</i> , 2004, 10 , 3479	Merck

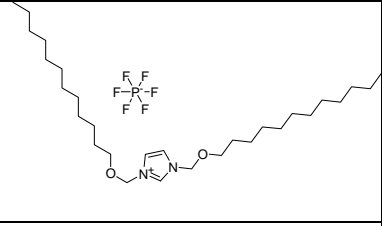
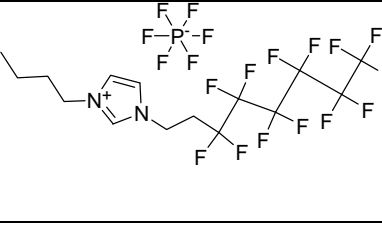
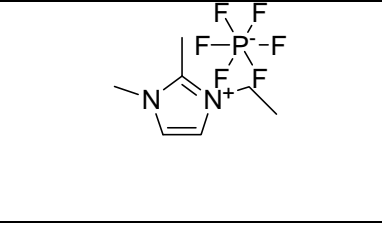
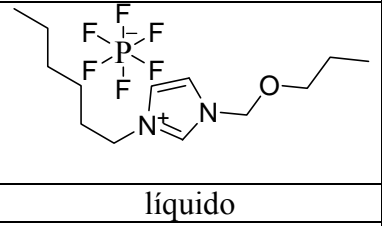
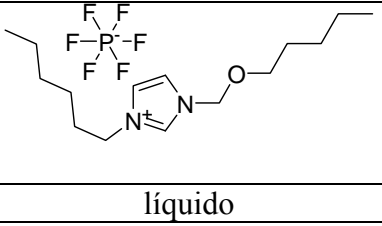
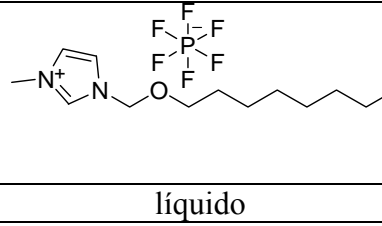
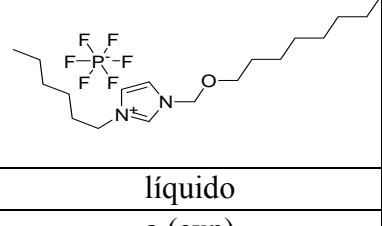
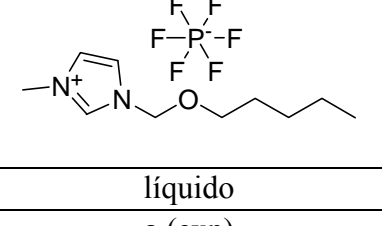
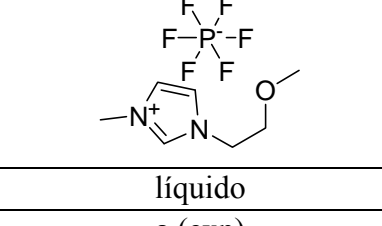
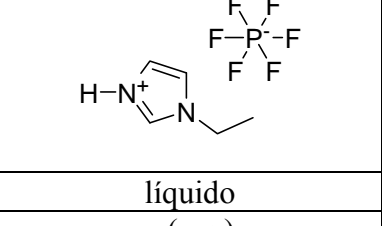
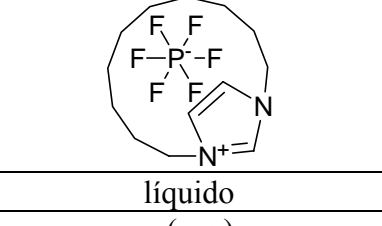
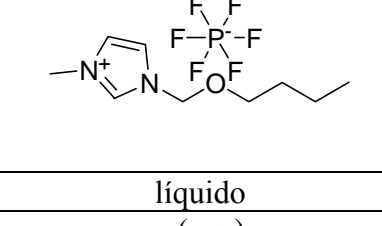
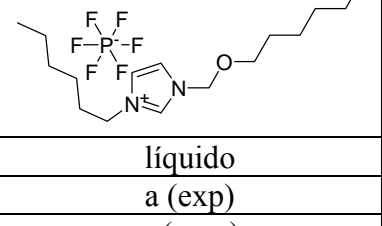
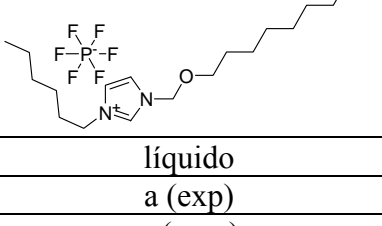
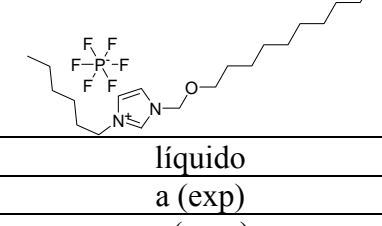
Capítulo II – Estudio de propiedades físico-químicas de líquidos iónicos

		
p.f. (exp): 49,1	p.f. (exp): 49,6	p.f. (exp): 55,2
b (exp)	b (exp)	b (exp)
b (prev)	b (prev)	b (prev)
Beilstein	Beilstein	Beilstein
		
p.f. (exp): 56,5	p.f. (exp): 56,5	p.f. (exp): 63
b (exp)	b (exp)	b (exp)
b (prev)	b (prev)	b (prev)
<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Chem. Eur. Journal</i> , 2004, 10 , 3479	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40
		
p.f. (exp): 66,8	p.f. (exp): 68,5	p.f. (exp): 79
b (exp)	b (exp)	b (exp)
b (prev)	b (prev)	b (prev)
<i>Dalton Trans.</i> , 1999, 2133	<i>Chem. Eur. Journal</i> , 2004, 10 , 3479	<i>Chem. Eur. Journal</i> , 2004, 10 , 3479
		
p.f. (exp): 85	líquido	líquido
b (exp)	a (exp)	a (exp)
b (prev)	a (prev)	a (prev)
<i>J. Am. Chem. Soc.</i> , 2004, 126 , 8114	<i>Inorg. Chem.</i> , 1996, 35 , 1168	Merck
		
líquido	líquido	líquido
a (exp)	a (exp)	a (exp)
a (prev)	a (prev)	a (prev)
<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40

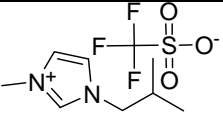
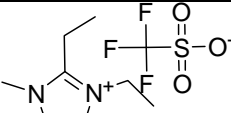
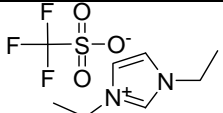
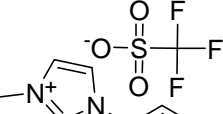
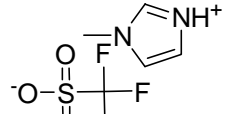
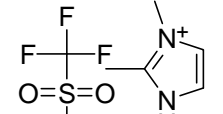
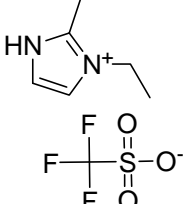
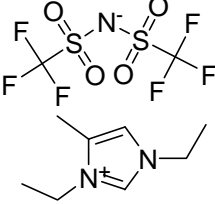
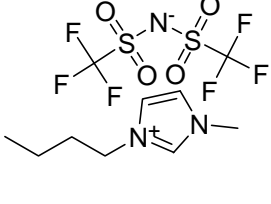
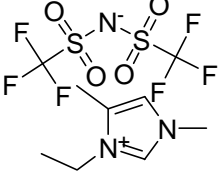
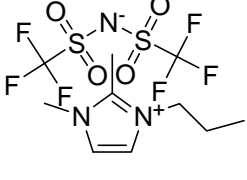
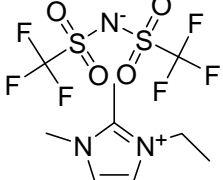
		
líquido	líquido	líquido
a (exp)	a (exp)	a (exp)
a (prev)	a (prev)	a (prev)
<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40
		
líquido	líquido	líquido
a (exp)	a (exp)	a (exp)
a (prev)	a (prev)	a (prev)
<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Org. Proc. Res. & Dev.</i> , 2002, 6 , 374
		
líquido	líquido	líquido
a (exp)	a (exp)	a (exp)
a (prev)	a (prev)	a (prev)
<i>Org. Proc. Res. & Dev.</i> , 2002, 6 , 374	<i>Chem. Eur. Journal</i> , 2004, 10 , 3479	<i>Chem. Eur. Journal</i> , 2004, 10 , 3479
		
líquido	líquido	líquido
a (exp)	a (exp)	a (exp)
a (prev)	a (prev)	a (prev)
<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Green Chem.</i> , 2005, 7 , 39

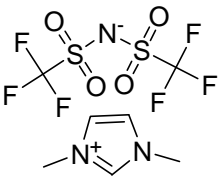
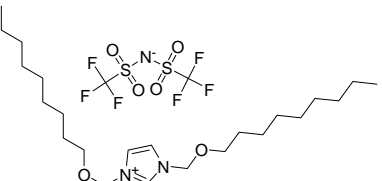
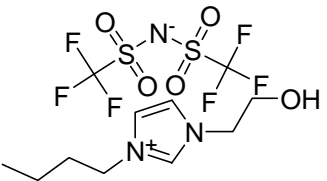
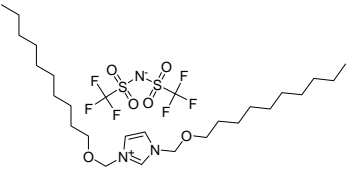
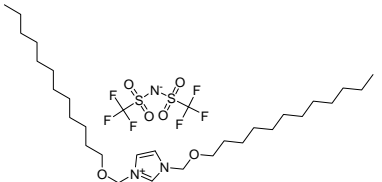
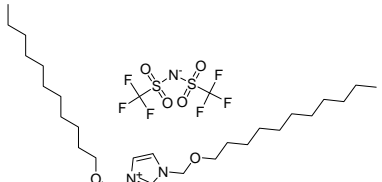
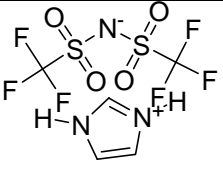
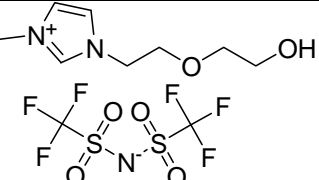
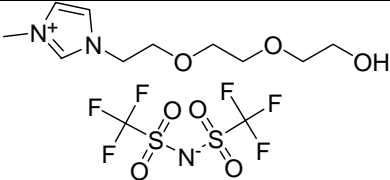
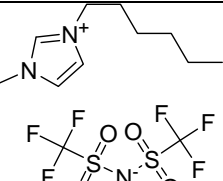
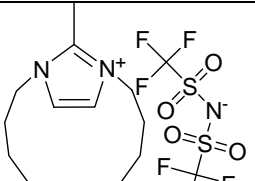
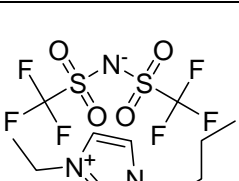
		
líquido	líquido	líquido
a (exp)	a (exp)	a (exp)
a (prev)	a (prev)	a (prev)
<i>J. Chrom. A</i> , 2004, 49 , 1037	<i>J. Chrom. A</i> , 2004, 49 , 1037	<i>Sol. State Ionics</i> ; 2002, 154 , 303
		
líquido	líquido	p.f. (exp): 52
a (exp)	a (exp)	b (exp)
a (prev)	a (prev)	b (prev)
<i>Sol. State Ionics</i> ; 2002, 154 , 303	<i>Sol. State Ionics</i> ; 2002, 154 , 303	<i>Chem. Eur. Journal</i> , 2004, 10 , 3479
		
p.f. (exp): 52,1	p.f. (exp): 5	p.f. (exp): 37,5
b (exp)	a (exp)	b (exp)
b (prev)	a (prev)	a (prev)
<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Green Chem.</i> , 2001, 3 , 156	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40
		
p.f. (exp): 38	p.f. (exp): 40	p.f. (exp): 48
b (exp)	b (exp)	b (exp)
b (prev)	b (prev)	b (prev)
<i>Ind. Chem. Res.</i> , 2001, 11 , 40	<i>Dalton Trans.</i> , 1999, 2133	<i>Ind. Chem. Res.</i> , 2001, 11 , 40
		
p.f. (exp): 48	p.f. (exp): 50	p.f. (exp): 53
b (exp)	b (exp)	b (exp)
b (prev)	b (prev)	b (prev)
<i>Ind. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Chem. Res.</i> , 2001, 11 , 40	<i>J. Org. Chem.</i> , 2002, 67 , 6845

p.f. (exp): 54,5	p.f. (exp): 57	p.f. (exp): 60,5
b (exp)	b (exp)	b (exp)
b (prev)	b (prev)	b (prev)
<i>Chem. Eur. Journal</i> , 2004, 10 , 3479	<i>Ind. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Chem. Res.</i> , 2001, 11 , 40
p.f. (exp): 61	p.f. (exp): 62	p.f. (exp): 70,7
b (exp)	b (exp)	b (exp)
b (prev)	b (prev)	b (prev)
<i>Chem. Eur. Journal</i> , 2004, 10 , 3479	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>J. Org. Chem.</i> , 2004, 69 , 6511
p.f. (exp): 71	p.f. (exp): 78,5	p.f. (exp): 85,5
b (exp)	b (exp)	b (exp)
b (prev)	b (prev)	b (prev)
<i>Chem. Eur. Journal</i> , 2004, 10 , 3479	<i>Chem. Eur. Journal</i> , 2004, 10 , 3479	<i>Chem. Eur. Journal</i> , 2004, 10 , 3479
p.f. (exp): 98	p.f. (exp): 102	p.f. (exp): 104
b (exp)	c (exp)	c (exp)
b (prev)	c (prev)	c (prev)
<i>Chem. Eur. Journal</i> , 2004, 10 , 3479	<i>Green Chem.</i> , 2001, 3 , 156	<i>Chem. Eur. Journal</i> , 2004, 10 , 3479

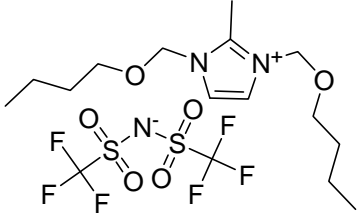
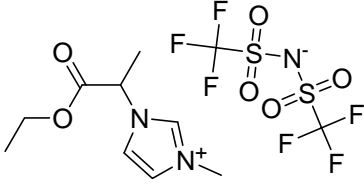
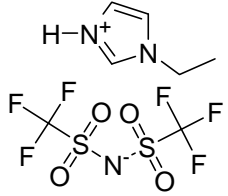
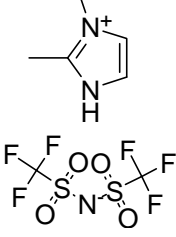
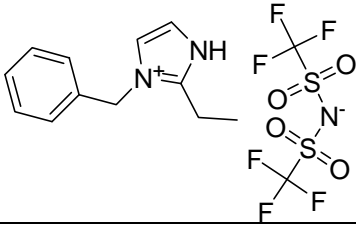
		
p.f. (exp): 105,5	p.f. (exp): 120	p.f. (exp): 196
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>Chem. Eur. Journal</i> , 2004, 10 , 3479	<i>Chem. Comm.</i> , 2000, 2051	Merck
		
líquido	líquido	líquido
a (exp)	a (exp)	a (exp)
a (prev)	a (prev)	a/b (prev)
<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40
		
líquido	líquido	líquido
a (exp)	a (exp)	a (exp)
a (prev)	a (prev)	a (prev)
<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Green Chem.</i> , 2005, 7 , 39
		
líquido	líquido	líquido
a (exp)	a (exp)	a (exp)
a (prev)	a (prev)	a (prev)
<i>Sol. State Ionics</i> , 2002, 154 , 303	<i>Chem. Comm.</i> , 2002, 2240	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40
		
líquido	líquido	líquido
a (exp)	a (exp)	a (exp)
a (prev)	a (prev)	a (prev)
<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40

líquido	líquido	líquido
a (exp)	a (exp)	a (exp)
a (prev)	a (prev)	a (prev)
<i>Org. Proc. Res. & Dev.</i> , 2002, 6 , 374	<i>Org. Proc. Res. & Dev.</i> , 2002, 6 , 374	<i>Org. Proc. Res. & Dev.</i> , 2002, 6 , 374
líquido	p.f. (exp): 115	líquido
a (exp)	c (exp)	a (exp)
a (prev)	c (prev)	a (prev)
<i>J. Chrom. A</i> , 2004, 49 , 1037	<i>Sol. State Ionics</i> , 2002, 154 , 303	<i>Sol. State Ionics</i> , 2002, 154 , 303
p.f. (exp): -10,5	p.f. (exp): 6	p.f. (exp): 16,5
a (exp)	a (exp)	a (exp)
a (prev)	a (prev)	a (prev)
Merck	<i>Inorg. Chem.</i> , 1996, 35 , 1168	<i>Green Chem.</i> , 2001, 3 , 156
p.f. (exp): 35	p.f. (exp): 38	p.f. (exp): 39
b (exp)	b (exp)	b (exp)
b (prev)	b (prev)	b (prev)
<i>Inorg. Chem.</i> , 1996, 35 , 1168	Merck	<i>Inorg. Chem.</i> , 1996, 35 , 1168
p.f. (exp): 45	p.f. (exp): 95	p.f. (exp): 191
b (exp)	b (exp)	c (exp)
b (prev)	b (prev)	c (prev)
<i>Inorg. Chem.</i> , 1996, 35 , 1168	Merck	<i>Sol. State Ionics</i> , 2002, 154 , 303

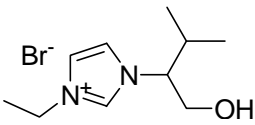
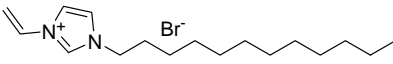
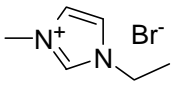
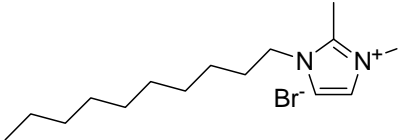
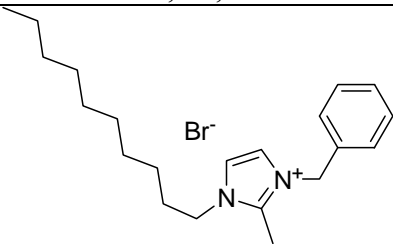
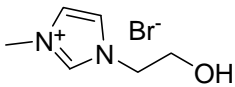
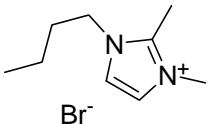
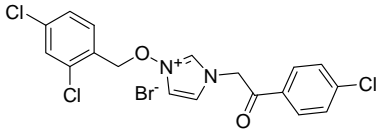
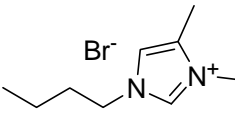
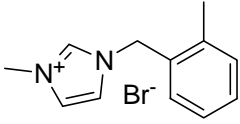
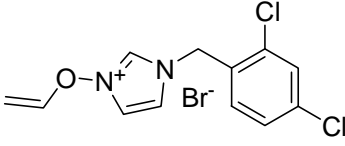
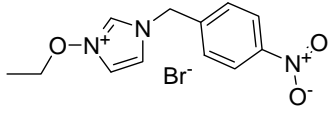
		
líquido	p.f. (exp): 113	líquido
a (exp)	c (exp)	a (exp)
a (prev)	c (prev)	a (prev)
<i>J. Chrom. A</i> , 2004, 49 , 1037	<i>Inorg. Chem.</i> , 1996, 35 , 1168	<i>Inorg. Chem.</i> , 1996, 35 , 1168
		
líquido	p.f. (exp): 84	p.f. (exp): 119
a (exp)	b (exp)	c (exp)
a (prev)	b (prev)	c (prev)
<i>Biomacromolecules</i> , 2004, 5 , 266	<i>Sol. State Ionics</i> , 2002, 154 , 303	<i>Sol. State Ionics</i> , 2002, 154 , 303
		
p.f. (exp): 33	p.f. (exp): -22	p.f. (exp): -4,9
b (exp)	a (exp)	a (exp)
b (prev)	a (prev)	a (prev)
<i>Sol. State Ionics</i> , 2002, 154 , 303	<i>Inorg. Chem.</i> , 1996, 35 , 1168	<i>Inorg. Chem.</i> , 1996, 35 , 1168
		
p.f. (exp): -3	p.f. (exp): 15	p.f. (exp): 20
a (exp)	a (exp)	a (exp)
a (prev)	a (prev)	a (prev)
<i>Inorg. Chem.</i> , 1996, 35 , 1168	Aldrich	<i>Inorg. Chem.</i> , 1996, 35 , 1168

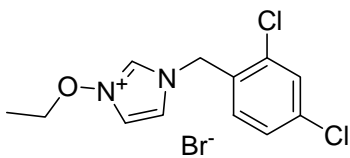
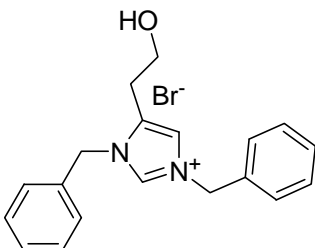
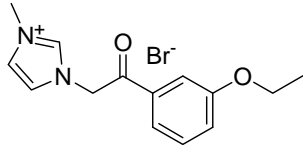
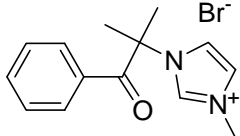
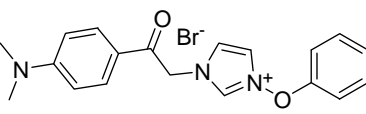
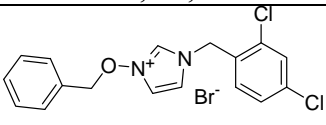
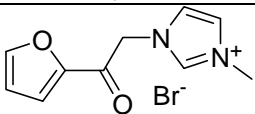
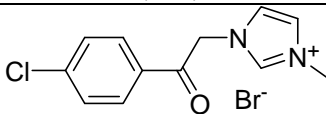
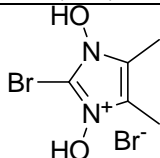
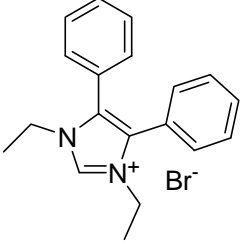
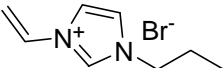
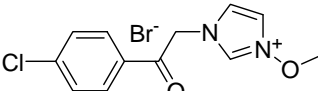
		
p.f. (exp): 22	p.f. (exp): 28	p.f. (exp): 30
a (exp)	b (exp)	b (exp)
a (prev)	b (prev)	a (prev)
<i>Inorg. Chem.</i> , 1996, 35 , 1168	<i>Chem. Eur. J.</i> , 2004, 10 , 3479	<i>Org. Proc. Res. Dev.</i> , 2002, 6 , 374
		
p.f. (exp): 31,5	p.f. (exp): 44,5	p.f. (exp): 45,5
b (exp)	b (exp)	b (exp)
b (prev)	b (prev)	b (prev)
<i>Chem. Eur. J.</i> , 2004, 10 , 3479	<i>Chem. Eur. J.</i> , 2004, 10 , 3479	<i>Chem. Eur. J.</i> , 2004, 10 , 3479
		
p.f. (exp): 73	líquido	líquido
b (exp)	a (exp)	a (exp)
b (prev)	a (prev)	a (prev)
<i>Chem. Comm.</i> , 2003, 938	<i>Org. Proc. Res. Dev.</i> , 2002, 6 , 374	<i>Org. Proc. Res. Dev.</i> , 2002, 6 , 374
		
líquido	p.f. (exp): 98	líquido
a (exp)	b (exp)	a (exp)
a (prev)	b (prev)	a (prev)
<i>Green Chem.</i> , 2002, 4 , 143	<i>Chem. Comm.</i> , 2002, 2240	<i>Inorg. Chem.</i> , 1996, 35 , 1168

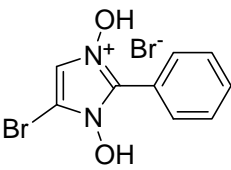
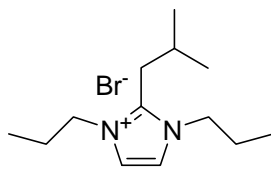
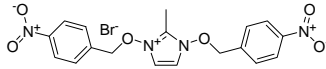
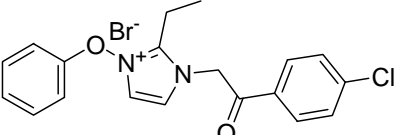
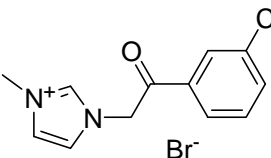
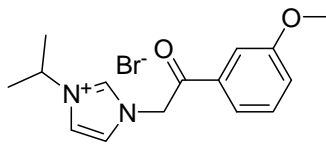
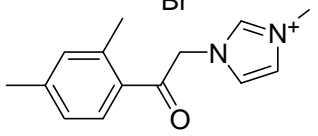
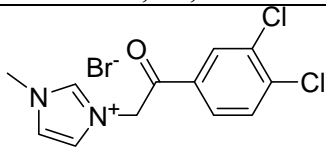
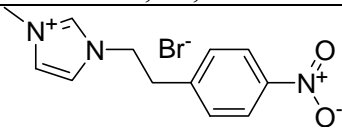
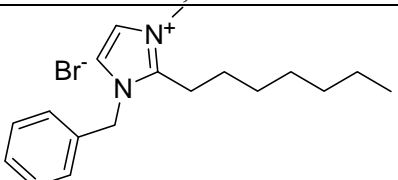
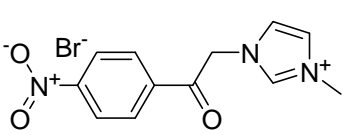
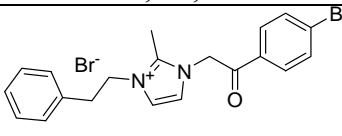
líquido	líquido	líquido
a (exp)	a (exp)	a (exp)
a (prev)	a (prev)	a (prev)
<i>Inorg. Chem.</i> , 1996, 35 , 1168	<i>J. Phys. Chem. B</i> , 2003, 42 , 107	<i>J. Org. Chem.</i> , 2004, 69 , 6511
líquido	líquido	líquido
a (exp)	a (exp)	a (exp)
a (prev)	a (prev)	a (prev)
<i>J. Org. Chem.</i> , 2004, 69 , 6511	<i>J. Org. Chem.</i> , 2004, 69 , 6511	<i>J. Org. Chem.</i> , 2004, 69 , 6511
líquido	líquido	líquido
a (exp)	a (exp)	a (exp)
a (prev)	a (prev)	a (prev)
<i>J. Org. Chem.</i> , 2004, 69 , 6511	<i>J. Org. Chem.</i> , 2004, 69 , 6511	<i>Chem. Eur. J.</i> , 2004, 10 , 3479
líquido	líquido	líquido
a (exp)	a (exp)	a (exp)
a (prev)	a (prev)	a (prev)
<i>Chem. Eur. J.</i> , 2004, 10 , 3479	<i>Chem. Eur. J.</i> , 2004, 10 , 3479	<i>J. Chrom. A</i> , 2004, 49 , 1037

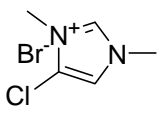
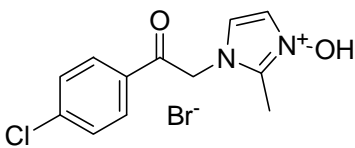
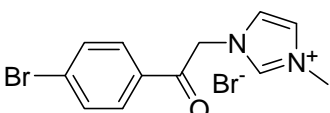
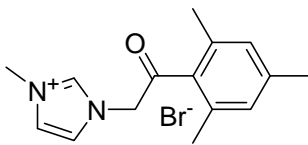
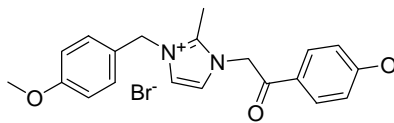
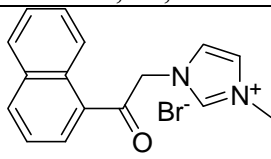
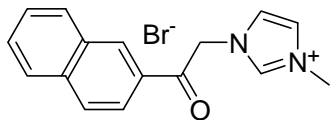
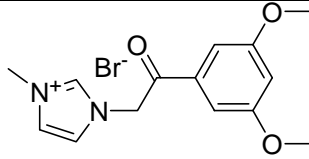
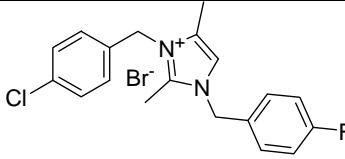
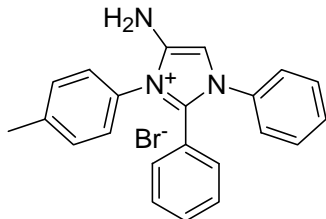
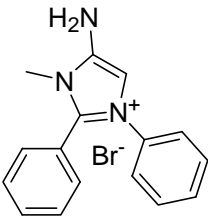
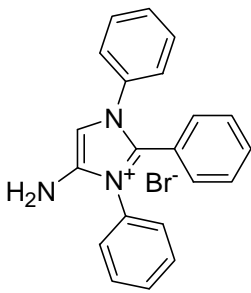
		
líquido	líquido	líquido
a (exp)	a (exp)	a (exp)
a (prev)	a (prev)	a (prev)
<i>Chem. Eur. J.</i> , 2004, 10 , 3479	<i>Tetrahedron Lett.</i> , 2004, 45 , 4429	<i>Sol. State Ionics</i> , 2002, 154 , 303
		
p.f. (exp): 22	líquido	
a (exp)	a (exp)	
a (prev)	a (prev)	
<i>Sol. State Ionics</i> , 2002, 154 , 303	<i>Sol. State Ionics</i> , 2002, 154 , 303	

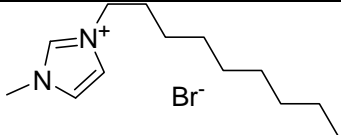
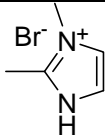
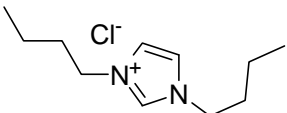
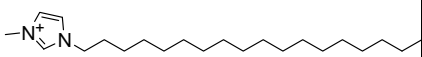
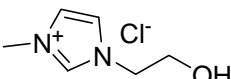
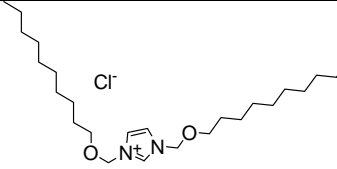
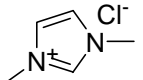
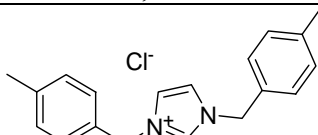
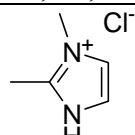
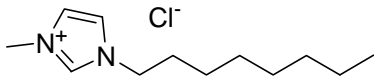
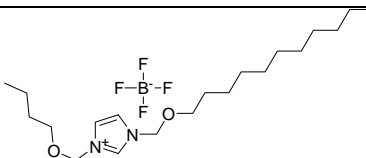
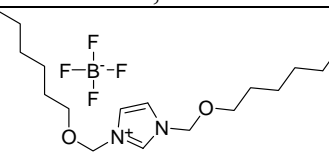
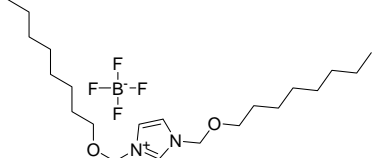
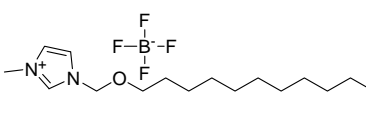
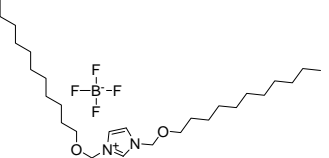
Conjunto de Teste:

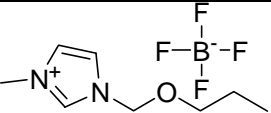
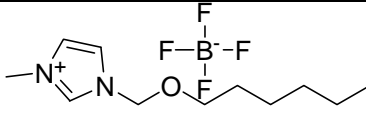
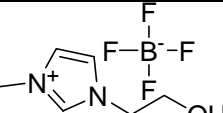
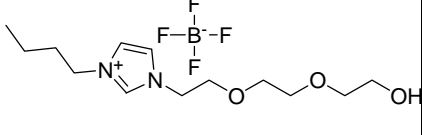
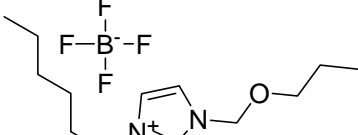
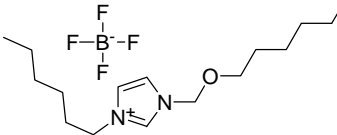
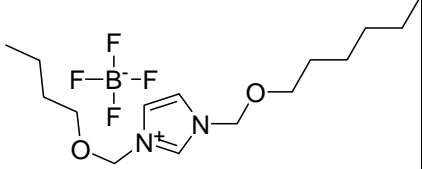
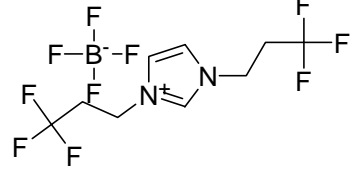
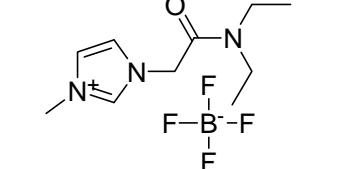
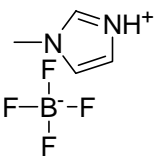
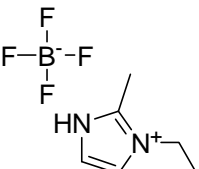
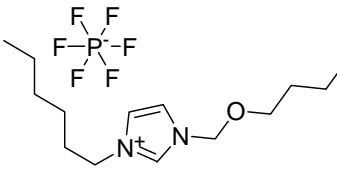
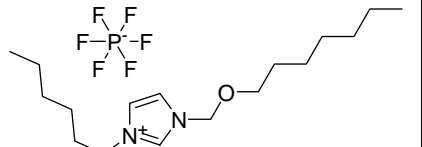
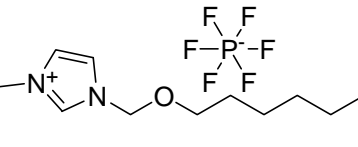
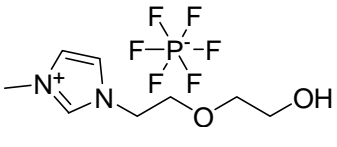
		
p.f. (exp): 12	p.f. (exp): 48	p.f. (exp): 74,5
a (exp)	b (exp)	b (exp)
a (prev)	b (prev)	b (prev)
<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225
		
p.f. (exp): 84,5	p.f. (exp): 87,5	p.f. (exp): 97,35
b (exp)	b (exp)	b (exp)
b (prev)	c (prev)	b (prev)
<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225
		
p.f. (exp): 100	p.f. (exp): 105	p.f. (exp): 110,5
b (exp)	c (exp)	c (exp)
b (prev)	c (prev)	b (prev)
<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225
		
p.f. (exp): 115	p.f. (exp): 117	p.f. (exp): 120
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225

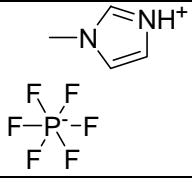
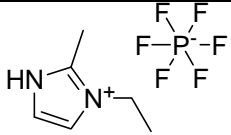
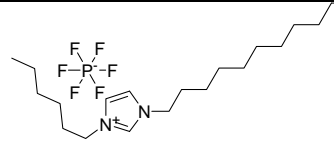
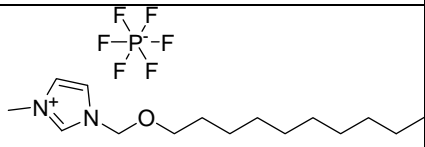
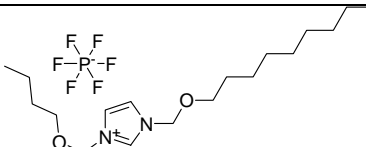
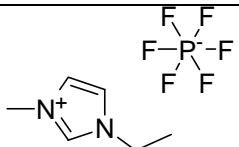
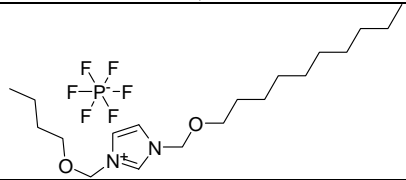
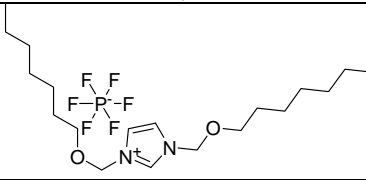
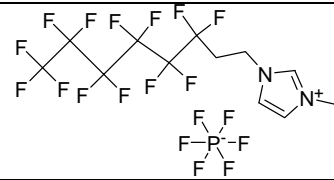
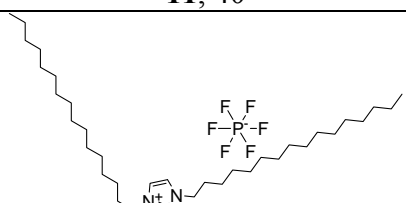
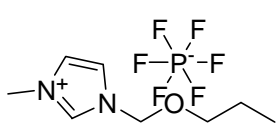
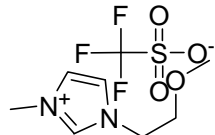
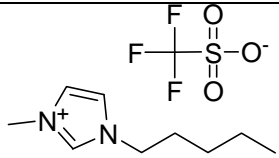
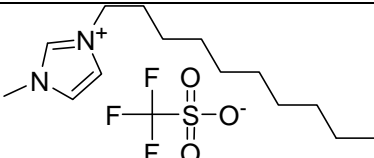
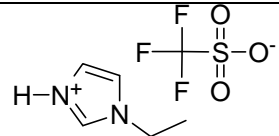
		
p.f. (exp): 134	p.f. (exp): 135	p.f. (exp): 137,5
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225
		
p.f. (exp): 138,5	p.f. (exp): 141	p.f. (exp): 144
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225
		
p.f. (exp): 147	p.f. (exp): 148,5	p.f. (exp): 152
c (exp)	c (exp)	c (exp)
a/c (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225
		
p.f. (exp): 153,5	p.f. (exp): 155	p.f. (exp): 156
c (exp)	c (exp)	c (exp)
c (prev)	b (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225

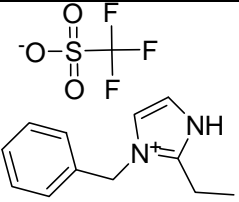
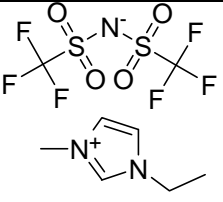
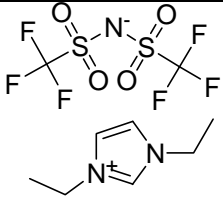
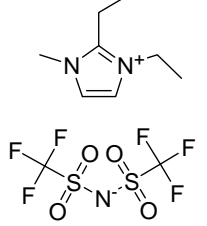
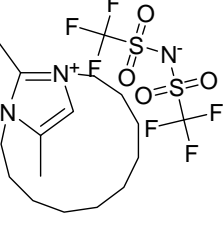
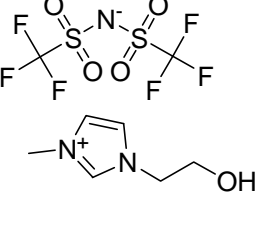
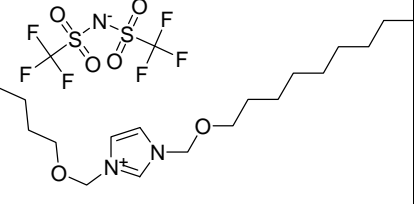
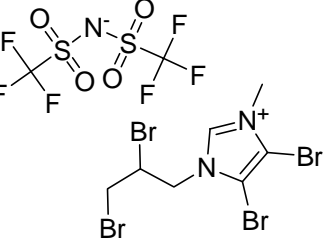
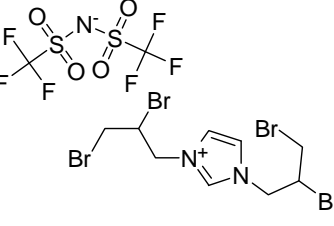
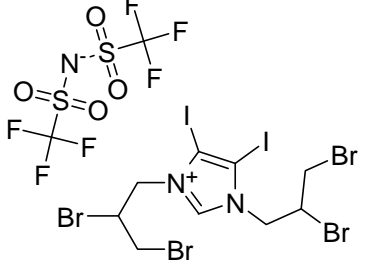
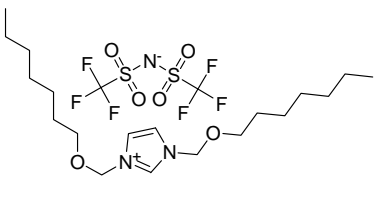
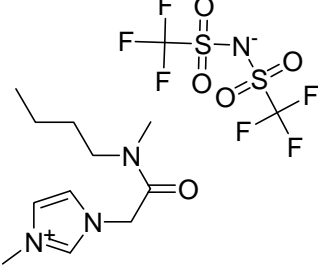
		
p.f. (exp): 158	p.f. (exp): 162,5	p.f. (exp): 168
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225
		
p.f. (exp): 171	p.f. (exp): 176	p.f. (exp): 178
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225
		
p.f. (exp): 180	p.f. (exp): 183,5	p.f. (exp): 185
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225
		
p.f. (exp): 186,5	p.f. (exp): 187,5	p.f. (exp): 193
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225

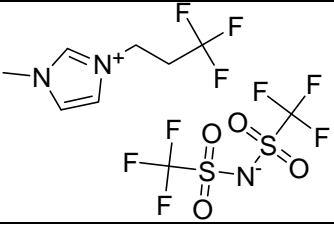
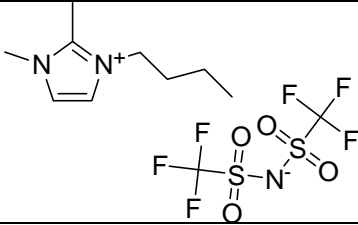
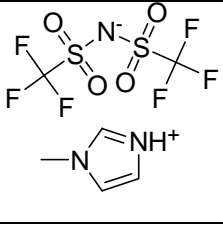
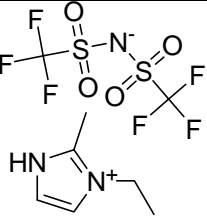
		
p.f. (exp): 195,5	p.f. (exp): 200	p.f. (exp): 218,5
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225
		
p.f. (exp): 222,5	p.f. (exp): 224,5	p.f. (exp): 230,5
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225
		
p.f. (exp): 238,5	p.f. (exp): 240	p.f. (exp): 250
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225
		
p.f. (exp): 267,5	p.f. (exp): 282,5	p.f. (exp): 283
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225

		
líquido	p.f. (exp): 176	p.f. (exp): 59,5
a (exp)	c (exp)	b (exp)
a (prev)	c (prev)	b (prev)
<i>J. Chem. Inf. Comp. Sci.</i> , 2002, 42 , 225	<i>Sol. State Ionics</i> , 2002, 154 , 303	<i>Synthesis</i> , 1996, 697
		
p.f. (exp): 70	p.f. (exp): 87	p.f. (exp): 100,5
b (exp)	b (exp)	c (exp)
c (prev)	b (prev)	b (prev)
Merck	<i>Org. Proc. Res. Dev.</i> , 2002, 6 , 374	<i>Chem. Eur. Journal</i> , 2004, 10 , 3479
		
p.f. (exp): 125	p.f. (exp): 157,5	p.f. (exp): 178
c (exp)	c (exp)	c (exp)
c (prev)	c (prev)	c (prev)
<i>Chem. Comm.</i> , 2001, 2399	<i>Synthesis</i> , 1996, 697	<i>Sol. State Ionics</i> , 2002, 154 , 303
		
p.f. (exp): -87	p.f. (exp): 22	p.f. (exp): 42,5
a (exp)	a (exp)	b (exp)
a (prev)	b (prev)	a (prev)
<i>Green Chem.</i> , 2001, 3 , 156	<i>Chem. Eur. J.</i> , 2004, 10 , 3479	<i>Chem. Eur. J.</i> , 2004, 10 , 3479
		
p.f. (exp): 53,5	p.f. (exp): 61,5	p.f. (exp): 71
b (exp)	b (exp)	b (exp)
b (prev)	b (prev)	b (prev)
<i>Chem. Eur. J.</i> , 2004, 10 , 3479	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Chem. Eur. Journal</i> , 2004, 10 , 3479

		
líquido	líquido	líquido
a (exp)	a (exp)	a (exp)
a (prev)	a (prev)	a (prev)
<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Org. Proc. Res. Dev.</i> , 2002, 6 , 374
		
líquido	líquido	líquido
a (exp)	a (exp)	a (exp)
a (prev)	a (prev)	a (prev)
<i>Chem Eur. J.</i> , 2002, 8 , 3671	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40
		
líquido	líquido	líquido
a (exp)	a (exp)	a (exp)
a (prev)	a (prev)	a/b (prev)
<i>Ind. Chem. Res.</i> , 2001, 11 , 40	<i>J. Chrom. A</i> , 2004, 49 , 1037	<i>Green Chem.</i> , 2004, 6 , 166
		
p.f. (exp): 37	líquido	líquido
b (exp)	a (exp)	a (exp)
a (prev)	a (prev)	a (prev)
<i>Sol. State Ionics</i> , 2002, 154 , 303	<i>Sol. State Ionics</i> , 2002, 154 , 303	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40
		
líquido	líquido	líquido
a (exp)	a (exp)	a (exp)
a (prev)	a (prev)	a (prev)
<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Org. Proc. Res. Dev.</i> , 2002, 6 , 374

		
p.f. (exp): 116	líquido	p.f. (exp): 71
c (exp)	a (exp)	b (exp)
a (prev)	c (prev)	a (prev)
<i>Sol. State Ionics</i> , 2002, 154 , 303	<i>Sol. State Ionics</i> , 2002, 154 , 303	<i>Green Chem.</i> , 2004, 6 , 166
		
p.f. (exp): 30	p.f. (exp): 50	p.f. (exp): 59,5
b (exp)	b (exp)	b (exp)
b (prev)	b (prev)	a (prev)
<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Green Chem.</i> , 2001, 3 , 156
		
p.f. (exp): 54	p.f. (exp): 64,5	p.f. (exp): 79,5
b (exp)	b (exp)	b (exp)
b (prev)	b (prev)	c (prev)
<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Chem. Eur. J.</i> , 2004, 10 , 3479	Aldrich
		
p.f. (exp): 103,3	líquido	p.f. (exp): 42
c (exp)	a (exp)	b (exp)
c (prev)	a (prev)	a (prev)
Beilstein	Beilstein	<i>Inorg. Chem.</i> , 1996, 35 , 1168
		
p.f. (exp): 109	líquido	p.f. (exp): 8
c (exp)	a (exp)	a (exp)
a (prev)	c (prev)	b (prev)
Merck	<i>J. Chrom. A</i> , 2004, 49 , 1037	<i>Sol. State Ionics</i> , 2002, 154 , 303

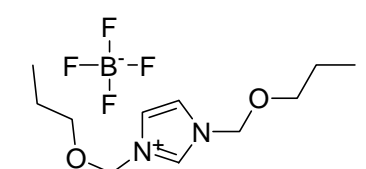
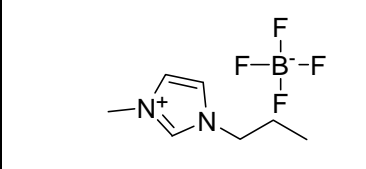
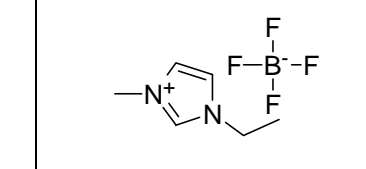
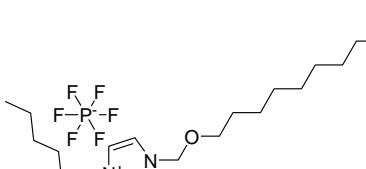
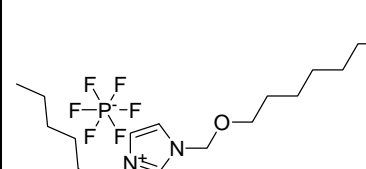
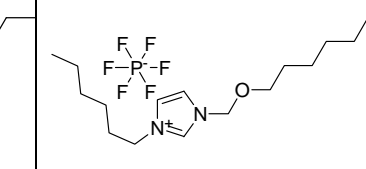
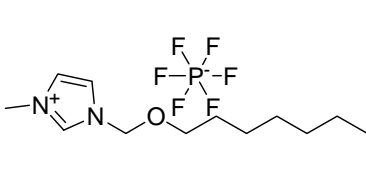
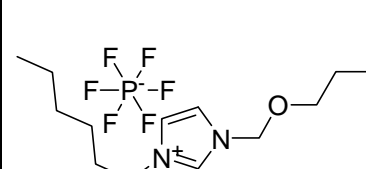
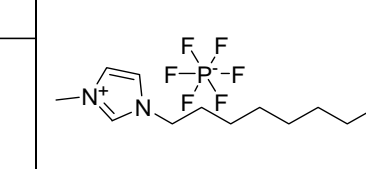
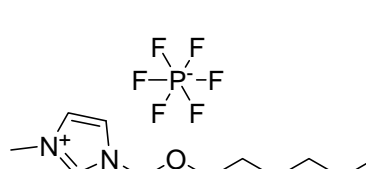
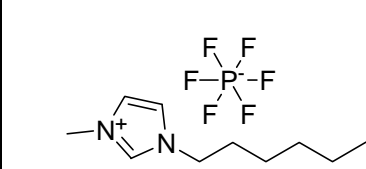
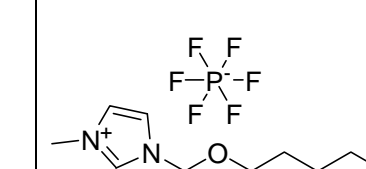
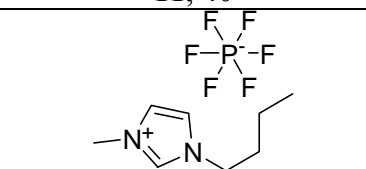
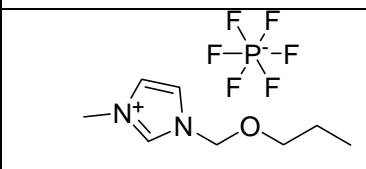
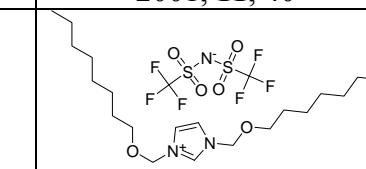
		
líquido	p.f. (exp): 14	p.f. (exp): 28
a (exp)	a (exp)	b (exp)
c (prev)	a (prev)	a (prev)
<i>Sol. State Ionics</i> , 2002, 154 , 303	<i>Inorg. Chem.</i> , 1996, 35 , 1168	<i>Inorg. Chem.</i> , 1996, 35 , 1168
		
p.f. (exp): 43,5	p.f. (exp): 86	líquido
b (exp)	b (exp)	a (exp)
a (prev)	b (prev)	a (prev)
<i>Inorg. Chem.</i> , 1996, 35 , 1168	<i>Chem. Comm.</i> , 2002, 2240	<i>Org. Proc. Res. Dev.</i> , 2002, 6 , 374
		
líquido	líquido	líquido
a (exp)	a (exp)	a (exp)
a (prev)	a (prev)	a (prev)
<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>J. Org. Chem.</i> , 2004, 69 , 6511	<i>J. Org. Chem.</i> , 2004, 69 , 6511
		
líquido	líquido	líquido
a (exp)	a (exp)	a (exp)
a (prev)	a (prev)	b (prev)
<i>J. Org. Chem.</i> , 2004, 69 , 6511	<i>Chem. Eur. J.</i> , 2004, 10 , 3479	<i>Green Chem.</i> , 2004, 6 , 166

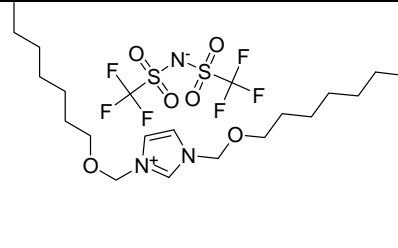
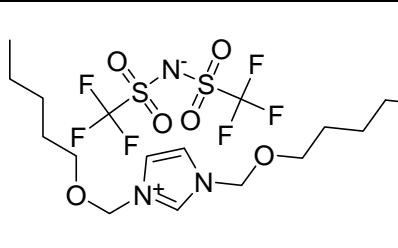
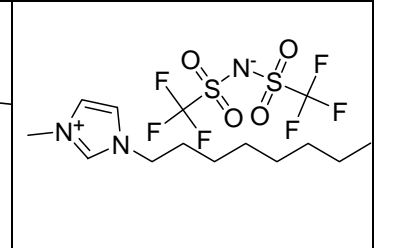
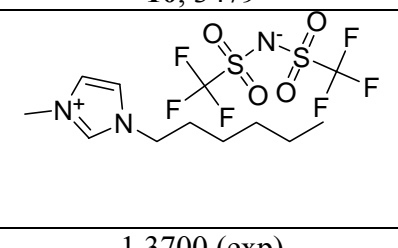
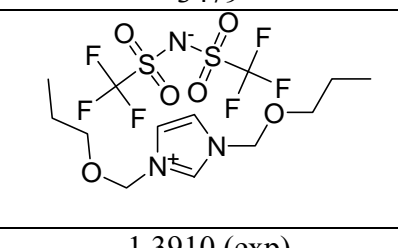
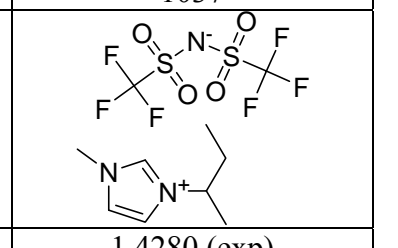
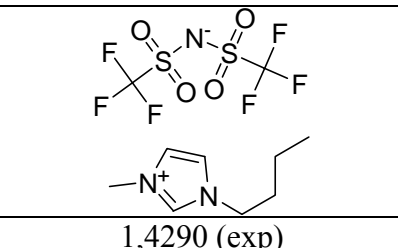
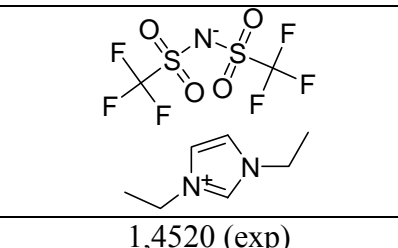
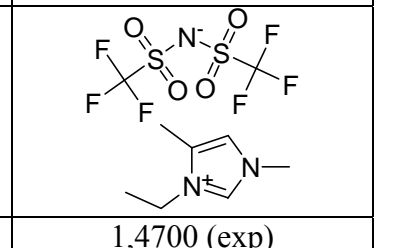
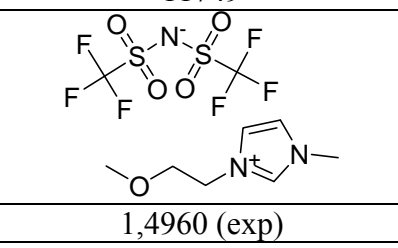
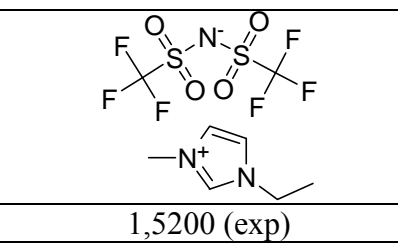
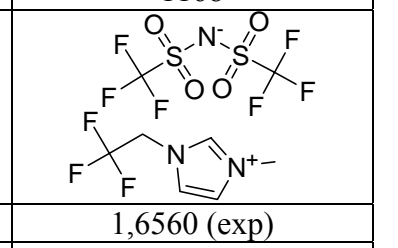
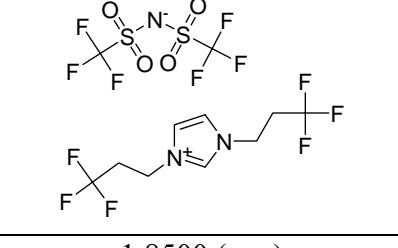
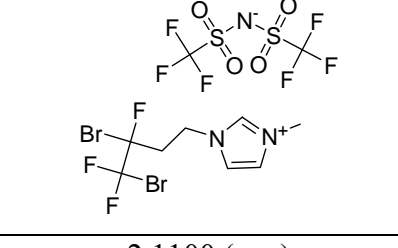
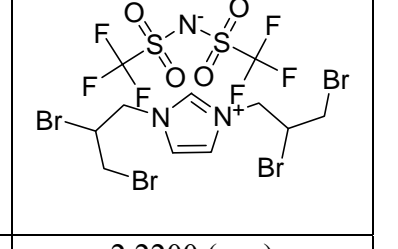
		
líquido	líquido	p.f. (exp): 9
a (exp)	a (exp)	a (exp)
a (prev)	a (prev)	a (prev)
<i>J. Chrom. A</i> , 2004, 49 , 1037	<i>J. Chrom. A</i> , 2004, 49 , 1037	<i>Sol. State Ionics</i> , 2002, 154 , 303
		
líquido		
a (exp)		
a (prev)		
<i>Sol. State Ionics</i> , 2002, 154 , 303		

Estudo c) : Modelação da densidade de sais de imidazólio (modelo SVM)

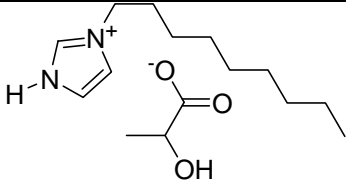
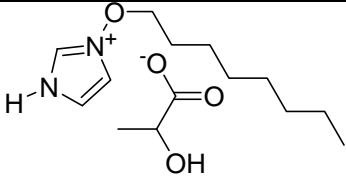
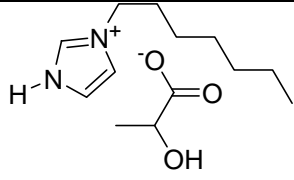
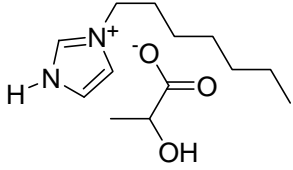
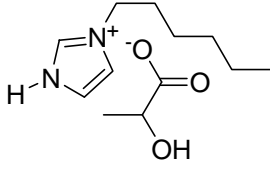
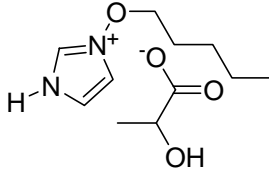
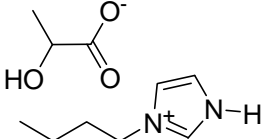
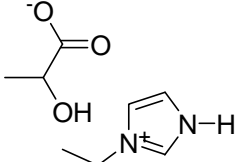
Conjunto de Treino

0,9900 (exp)	1,0000 (exp)	1,0300 (exp)
0,998467 (prev)	0,991901 (prev)	1,0342611 (prev)
<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40
1,0400 (exp)	1,0500 (exp)	1,0600 (exp)
1,029093 (prev)	1,048407 (prev)	1,063955 (prev)
<i>Ind. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40
1,0800 (exp)	1,0800 (exp)	1,1050 (exp)
1,089172 (prev)	1,076116 (prev)	1,094203 (prev)
<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40
1,1050 (exp)	1,1100 (exp)	1,1100 (exp)
1,094485 (prev)	1,102579 (prev)	1,09943 (prev)
<i>Ind. Chem. Res.</i> , 2001, 11 , 40	<i>J. Chrom. A</i> , 2004, 49 , 1037	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40
1,1500 (exp)	1,1500 (exp)	1,1900 (exp)
1,15888 (prev)	1,139149 (prev)	1,184541 (prev)
<i>J. Chrom. A</i> , 2004, 49 , 1037	<i>Chem. Eur. Journal</i> , 2004, 10 , 3479	<i>Ind. Chem. Res.</i> , 2001, 11 , 40

		
1,1940 (exp)	1,2400 (prev)	1,3400 (prev)
1,204789 (prev)	1,250896 (exp)	1,329839 (exp)
<i>Chem. Eur. Journal</i> , 2004, 10 , 3479	<i>Dalton .Trans.</i> , 1999, 2133	<i>Chem. Eur. J.</i> , 2004, 10 , 6581
		
1,0900 (exp)	1,1000 (exp)	1,1500 (exp)
1,099515 (prev)	1,099869 (prev)	1,155069 (prev)
<i>Ind. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Chem. Res.</i> , 2001, 11 , 40
		
1,1900 (exp)	1,2000 (exp)	1,2300 (exp)
1,200587 (prev)	1,190331 (prev)	1,230275 (prev)
<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Green Chem.</i> , 2001, 3 , 156
		
1,2650 (exp)	1,2950 (exp)	1,3300 (exp)
1,276084 (prev)	1,303347 (prev)	1,318983 (prev)
<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Green Chem.</i> , 2002, 4 , 143	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40
		
1,3600 (exp)	1,4100 (exp)	1,2050 (exp)
1,370134 (prev)	1,418998 (prev)	1,215804 (prev)
<i>Green Chem.</i> , 2001, 3 , 156	<i>Ind. Eng. Chem. Res.</i> , 2001, 11 , 40	<i>Chem. Eur. Journal</i> , 2004, 10 , 3479

		
1,2260 (exp)	1,2770 (exp)	1,3300 (exp)
1,222352 (prev)	1,287828 (prev)	1,319017 (prev)
<i>Chem. Eur. Journal</i> , 2004, 10 , 3479	<i>Chem. Eur. Journal</i> , 2004, 10 , 3479	<i>J. Chrom. A</i> , 2004, 49 , 1037
		
1,3700 (exp)	1,3910 (exp)	1,4280 (exp)
1,367718 (prev)	1,401959 (prev)	1,438953 (prev)
<i>Green Chem.</i> , 2002, 4 , 143	<i>Chem. Eur. Journal</i> , 2004, 10 , 3479	<i>J. Chrom. A</i> , 2004, 49 , 1037
		
1,4290 (exp)	1,4520 (exp)	1,4700 (exp)
1,427348 (prev)	1,462964 (prev)	1,480838 (prev)
<i>J. Phys. Chem. B</i> , 2003, 107 , 11749	<i>Inorg. Chem.</i> , 1996, 35 , 1168	<i>Inorg. Chem.</i> , 1996, 35 , 1168
		
1,4960 (exp)	1,5200 (exp)	1,6560 (exp)
1,485445 (prev)	1,5097 (prev)	1,646551 (prev)
<i>Inorg. Chem.</i> , 1996, 35 , 1168	<i>Inorg. Chem.</i> , 1996, 35 , 1168	<i>J. Chrom. A</i> , 2004, 49 , 1037
		
1,8500 (exp)	2,1100 (exp)	2,2200 (exp)
1,839032 (prev)	2,09902 (prev)	2,209041 (prev)
<i>J. Chrom. A</i> , 2004, 49 , 1037	<i>J. Org. Chem.</i> , 2004, 69 , 6511	<i>J. Org. Chem.</i> , 2004, 69 , 6511

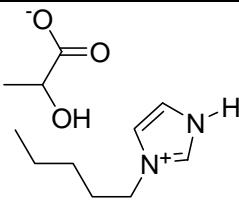
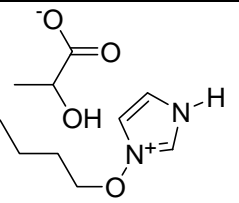
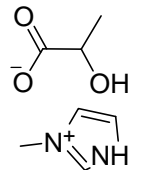
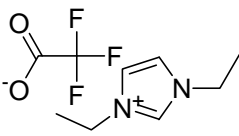
2,2600 (exp)	2,3000 (exp)	2,5500 (exp)
2,255681 (prev)	2,299226 (prev)	2,55001 (prev)
<i>J. Org. Chem.</i> , 2004, 69 , 6511	<i>J. Org. Chem.</i> , 2004, 69 , 6511	<i>J. Org. Chem.</i> , 2004, 69 , 6511
2,8000 (exp)	1,2400 (exp)	1,3300 (exp)
2,78905 (prev)	1,232089 (prev)	1,319014 (prev)
<i>J. Org. Chem.</i> , 2004, 69 , 6511	<i>Green Chem.</i> , 2002, 4 , 143	<i>Inorg. Chem.</i> , 1996, 35 , 1168
1,3900 (exp)	1,1830 (exp)	1,2150 (exp)
1,400824 (prev)	1,194155 (prev)	1,204505 (prev)
<i>Inorg. Chem.</i> , 1996, 35 , 1168	<i>Inorg. Chem.</i> , 1996, 35 , 1168	<i>Inorg. Chem.</i> , 1996, 35 , 1168
1,2900 (exp)	0,9591 (exp)	0,9685 (exp)
1,287577 (prev)	0,969399 (prev)	0,978477 (prev)
<i>Inorg. Chem.</i> , 1996, 35 , 1168	<i>Green Chem</i> , 2004., 6 , 323	<i>Green Chem.</i> , 2004, 6 , 323
0,9804 (exp)	0,9854 (exp)	0,9906 (exp)
0,990053 (prev)	0,975593 (prev)	0,995564 (prev)
<i>Green Chem.</i> , 2004, 6 , 323	<i>Green Chem.</i> , 2004, 6 , 323	<i>Green Chem.</i> , 2004, 6 , 323

		
0,9908 (exp)	1,0105 (exp)	1,0207 (exp)
0,982094 (prev)	1,000744 (prev)	1,018229 (prev)
<i>Green Chem.</i> , 2004, 6 , 323	<i>Green Chem.</i> , 2004, 6 , 323	<i>Green Chem.</i> , 2004, 6 , 323
		
1,0273 (exp)	1,0374 (exp)	1,0490 (exp)
1,017397 (prev)	1,034585 (prev)	1,060066 (prev)
<i>Green Chem.</i> , 2004, 6 , 323	<i>Green Chem.</i> , 2004, 6 , 323	<i>Green Chem.</i> , 2004, 6 , 323
		
1,0595 (exp)	1,1081 (exp)	
1,070441 (prev)	1,115018 (prev)	
<i>Green Chem.</i> , 2004, 6 , 323	<i>Green Chem.</i> , 2004, 6 , 323	

Conjunto de Teste

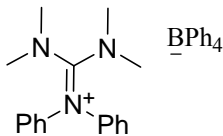
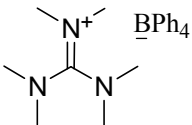
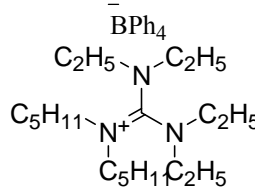
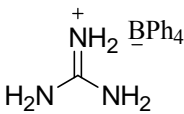
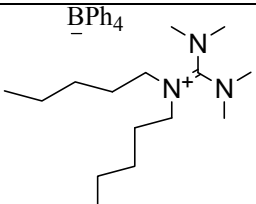
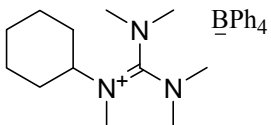
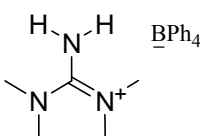
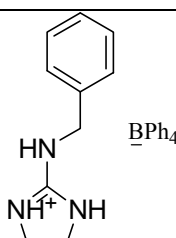
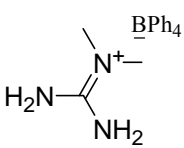
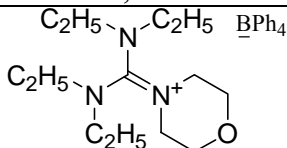
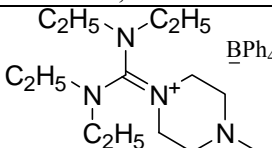
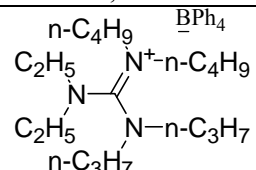
1,0200 (exp)	1,0480 (exp)	1,0650 (exp)
1,014203 (prev)	1,043101 (prev)	1,063544 (prev)
<i>Ind. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Chem. Res.</i> , 2001, 11 , 40
1,0950 (exp)	1,1050 (exp)	1,1150 (exp)
1,072644 (prev)	1,067554 (prev)	1,122912 (prev)
<i>Ind. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Chem. Res.</i> , 2001, 11 , 40
1,1800 (exp)	1,2000 (exp)	1,2050 (exp)
1,223952 (prev)	1,238766 (prev)	1,241158 (prev)
<i>Green Chem.</i> , 2001, 3 , 156	<i>Ind. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Chem. Res.</i> , 2001, 11 , 40
1,1250 (exp)	1,1750 (exp)	1,2250 (exp)
1,14155 (prev)	1,167322 (prev)	1,225577 (prev)
<i>Ind. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Chem. Res.</i> , 2001, 11 , 40
1,2750 (exp)	1,3350 (exp)	1,2490 (exp)
1,243428 (prev)	1,356936 (prev)	1,243729 (prev)
<i>Ind. Chem. Res.</i> , 2001, 11 , 40	<i>Ind. Chem. Res.</i> , 2001, 11 , 40	<i>Chem. Eur. Journal</i> , 2004, 10 , 3479

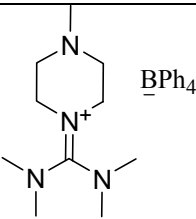
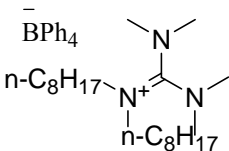
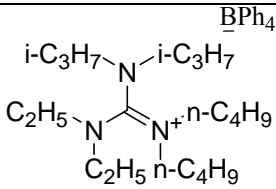
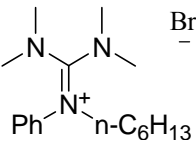
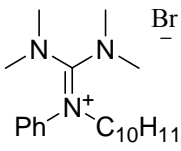
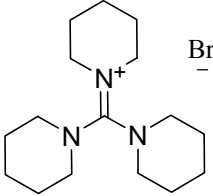
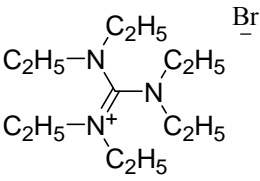
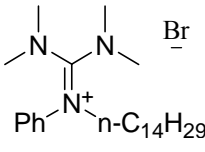
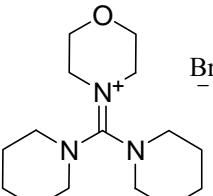
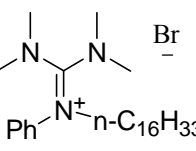
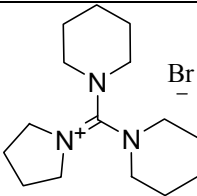
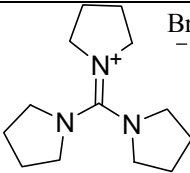
1,3380 (exp)	1,4040 (exp)	1,4320 (exp)
1,346073 (prev)	1,397403 (prev)	1,457438 (prev)
<i>Chem. Eur. Journal</i> , 2004, 10 , 3479	<i>Inorg. Chem.</i> , 1996, 35 , 1168	<i>Inorg. Chem.</i> , 1996, 35 , 1168
1,4950 (exp)	1,5590 (exp)	1,9500 (exp)
1,460144 (prev)	1,561805 (prev)	1,80497 (prev)
<i>Inorg. Chem.</i> , 1996, 35 , 1168	<i>Inorg. Chem.</i> , 1996, 35 , 1168	<i>J. Org. Chem.</i> , 2004, 69 , 6511
2,2500 (exp)	2,3500 (exp)	1,2700 (exp)
2,255929 (prev)	2,08457 (prev)	1,265457 (prev)
<i>J. Org. Chem.</i> , 2004, 69 , 6511	<i>J. Org. Chem.</i> , 2004, 69 , 6511	<i>Inorg. Chem.</i> , 1996, 35 , 1168
1,3340 (exp)	1,3640 (exp)	0,9793 (exp)
1,317311 (prev)	1,365451 (prev)	0,984086 (prev)
<i>Inorg. Chem.</i> , 1996, 35 , 1168	<i>Inorg. Chem.</i> , 1996, 35 , 1168	<i>Green Chem.</i> , 2004, 6 , 323
0,9861 (exp)	1,0026 (exp)	1,0252 (exp)
0,98423 (prev)	0,99171 (prev)	1,04158 (prev)
<i>Green Chem.</i> , 2004, 6 , 323	<i>Green Chem.</i> , 2004, 6 , 323	<i>Green Chem.</i> , 2004, 6 , 323

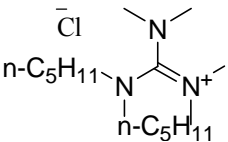
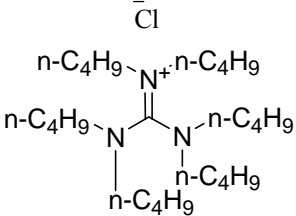
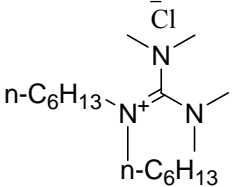
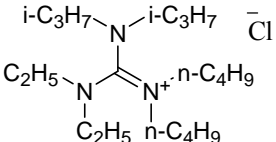
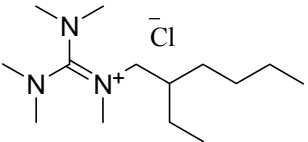
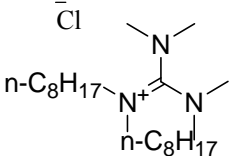
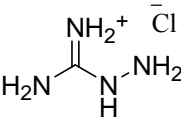
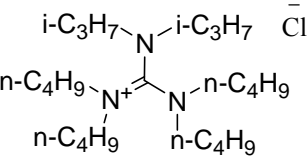
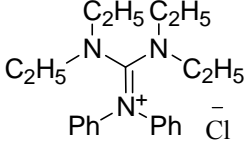
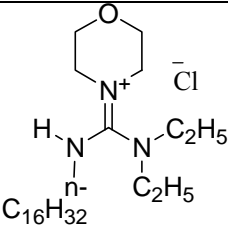
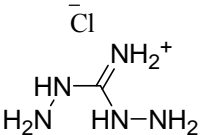
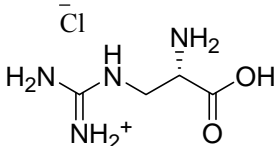
		
1,0413 (exp)	1,0640 (exp)	1,1231 (exp)
1,036749 (prev)	1,101246 (prev)	1,237867 (pred)
<i>Green Chem.</i> , 2004, 6 , 323	<i>Green Chem.</i> , 2004, 6 , 323	<i>Green Chem.</i> , 2004, 6 , 323
		
1,2500 (exp)		
1,259302 (prev)		
<i>Inorg. Chem.</i> , 1996, 35 , 1168		

Estudo d) : Modelação do ponto de fusão de sais de guanidínio (Rede CPG)

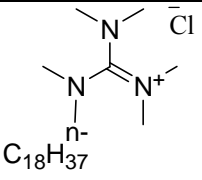
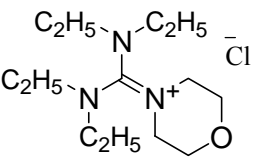
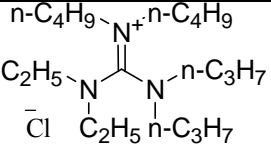
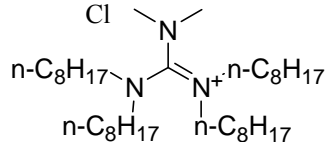
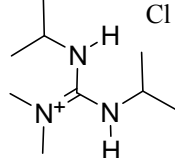
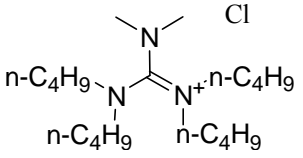
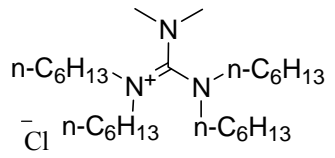
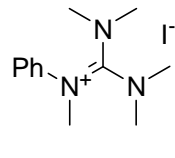
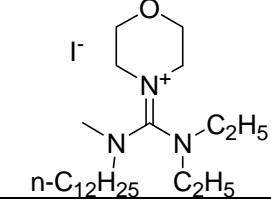
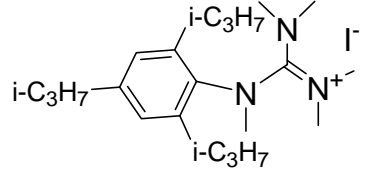
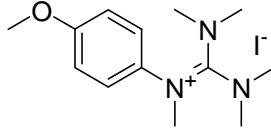
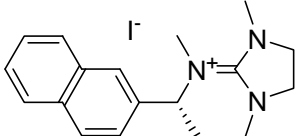
Conjunto de Treino

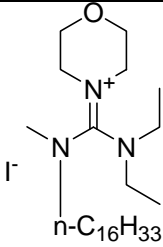
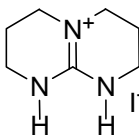
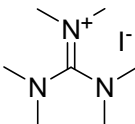
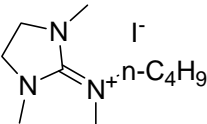
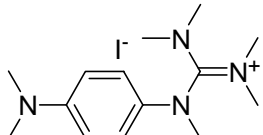
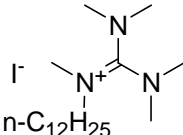
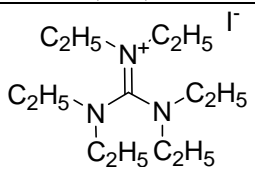
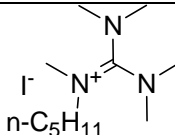
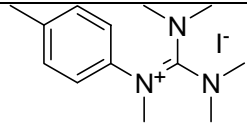
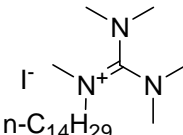
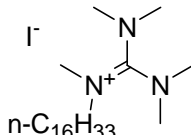
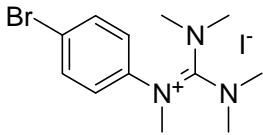
		
163°C (exp)	322°C (exp)	127°C (exp)
164,2°C (prev)	281,1°C (prev)	178,2°C (prev)
<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Liebigs Ann. Chem.</i> , 1984, 108
		
195°C (exp)	163°C (exp)	224°C (exp)
204,3°C (prev)	182,4°C (prev)	217,8°C (prev)
<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Liebigs Ann. Chem.</i> , 1984, 108
		
208°C (exp)	139°C (exp)	213°C (exp)
225,1°C (prev)	141,9°C (prev)	206,4°C (prev)
<i>Tetrahedron</i> , 1999, 55, 6027	<i>Tetrahedron</i> , 1999, 55, 6027	<i>Tetrahedron</i> , 1999, 55, 6027
		
272°C (exp)	210°C (exp)	203°C (exp)
227,6°C (prev)	227,6°C (prev)	179,8°C (prev)
<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Liebigs Ann. Chem.</i> , 1984, 108

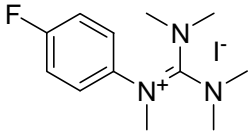
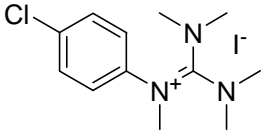
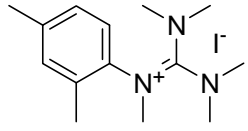
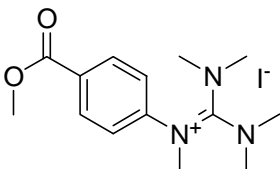
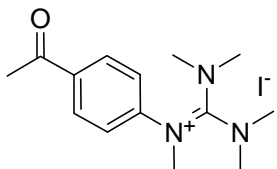
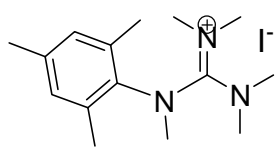
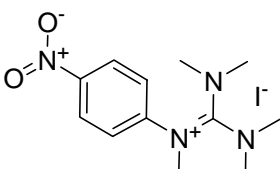
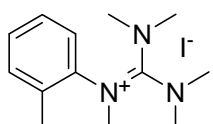
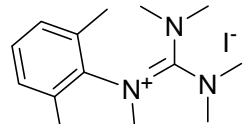
		
203°C (exp)	135°C (exp)	203°C (exp)
223°C (prev)	142,5°C (prev)	187,4°C (prev)
<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Liebigs Ann. Chem.</i> , 1984, 108
		
99°C (exp)	93°C (exp)	187°C (exp)
99,5°C (prev)	99,8°C (prev)	171,5°C (prev)
<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Liebigs Ann. Chem.</i> , 1984, 108	Patente Americana US5132423, 1992
		
166°C (exp)	103°C (exp)	166°C (exp)
169,8°C (prev)	103°C (prev)	171,5°C (prev)
<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Liebigs Ann. Chem.</i> , 1984, 108	Patente Americana US5132423, 1992
		
100°C (exp)	147°C (exp)	214°C (exp)
103°C (prev)	170,3°C (prev)	194,5°C (prev)
<i>Liebigs Ann. Chem.</i> , 1984, 108	Patente Americana US5132423, 1992	Patente Americana US5132423, 1992

		
69°C (exp)	147°C (exp)	95°C (exp)
63,9°C (prev)	100,9°C (prev)	58,9°C (prev)
<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Liebigs Ann. Chem.</i> , 1984, 108
		
55°C (exp)	142°C (exp)	64°C (exp)
38,6°C (prev)	132,9°C (prev)	37°C (prev)
<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Liebigs Ann. Chem.</i> , 1984, 108
		
164°C (exp)	85°C (exp)	50°C (exp)
182,3°C (prev)	96,6°C (prev)	101,6°C (prev)
Aldrich web site: http://www.sigmaaldrich.com	Beilstein	<i>Liebigs Ann. Chem.</i> , 1984, 108
		
71°C (exp)	181°C (exp)	220°C (exp)
76,6°C (prev)	182,3°C (prev)	199,9°C (prev)
<i>Liebigs Ann. Chem.</i> , 1984, 108	Aldrich web site: http://www.sigmaaldrich.com	Aldrich web site: http://www.sigmaaldrich.com

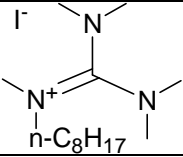
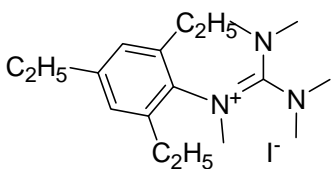
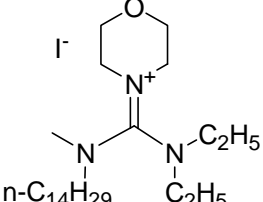
85°C (exp)	266°C (exp)	63°C (exp)
94,4°C (prev)	217,9°C (prev)	63,9°C (prev)
<i>Liebigs Ann. Chem.</i> , 1984, 108	Patente Alemã DE2716477, 1978	<i>Liebigs Ann. Chem.</i> , 1984, 108
58°C (exp)	293°C (exp)	50°C (exp)
11,1°C (prev)	284,7°C (prev)	44,3°C (prev)
<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Liebigs Ann. Chem.</i> , 1984, 108
53°C (exp)	59°C (exp)	72°C (exp)
69,4°C (prev)	82,8 (prev)	77,5°C (prev)
<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Liebigs Ann. Chem.</i> , 1984, 108
148°C (exp)	51°C (exp)	111°C (exp)
128,4°C (prev)	47,1°C (prev)	111,2°C (prev)
<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Liebigs Ann. Chem.</i> , 1984, 108

		
138°C (exp)	110,5°C (exp)	56°C (exp)
118,4°C (prev)	109,6°C (prev)	42,5°C (prev)
<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Liebigs Ann. Chem.</i> , 1984, 108
		
-76°C (exp)	292°C (exp)	-55°C (exp)
-49,6°C (prev)	162°C (prev)	11,06°C (prev)
<i>Green Chem.</i> , 2005, 7, 362	Patente Alemã DE2716477, 1978	<i>Green Chem.</i> , 2005, 7, 362
		
-63°C (exp)	133°C (exp)	52°C (exp)
-39,5°C (prev)	197,8°C (prev)	64,3°C (prev)
<i>Green Chem.</i> , 2005, 7, 362	<i>Chem. Ber</i> , 1971, 104 , 2158	<i>Liebigs Ann. Chem.</i> , 1984, 108
		
273°C (exp)	172°C (exp)	192°C (exp)
197,8°C (prev)	170,4°C (prev)	179,5°C (prev)
<i>J.Org Chem.</i> , 2000, 65 , 7770	<i>J.Org Chem.</i> , 2000, 65 , 7770	<i>J.Org Chem.</i> , 2000, 65 , 7770

		
70°C (exp)	118°C (exp)	300°C (exp)
108,6°C (prev)	116°C (prev)	293,6°C (prev)
<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Tetrahedron</i> , 1999, 55 , 6027	Patente Alemã DE2716477, 1978
		
46°C (exp)	104°C (exp)	196°C (exp)
89°C (prev)	151,8°C (prev)	192,4°C (prev)
<i>Tetrahedron Letters</i> , 2004, 45 , 2013	<i>Chem. Ber.</i> , 1971, 104 , 2158	<i>Liebigs Ann. Chem.</i> , 1984, 108
		
275°C (exp)	145°C (exp)	126°C (exp)
276,4°C (prev)	127,5°C (prev)	169°C (prev)
<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Chem. Ber.</i> , 1971, 104 , 2158
		
208°C (exp)	215°C (exp)	169°C (exp)
191,3°C (prev)	196,3°C (prev)	170,4°C (prev)
<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Chem. Ber.</i> , 1971, 104 , 2158

		
192°C (exp)	154°C (exp)	159°C (exp)
170,4°C (prev)	170,4°C (prev)	173,6°C (prev)
<i>Chem. Ber</i> , 1971, 104 , 2158	<i>Chem. Ber</i> , 1971, 104 , 2158	<i>Chem. Ber</i> , 1971, 104 , 2158
		
167°C (exp)	164°C (exp)	171°C (exp)
170,8°C (prev)	170,8°C (prev)	177,7°C (prev)
<i>Chem. Ber</i> , 1971, 104 , 2158	<i>Chem. Ber</i> , 1971, 104 , 2158	<i>Chem. Ber</i> , 1971, 104 , 2158
		
178°C (exp)	168°C (exp)	245°C (exp)
170,8°C (prev)	173,6°C (prev)	178,4°C (prev)
<i>Chem. Ber</i> , 1971, 104 , 2158	<i>Chem. Ber</i> , 1971, 104 , 2158	<i>Chem. Ber</i> , 1971, 104 , 2158

298

Erro! Marcador não definido.	Erro! Marcador não definido.	
99°C (exp)	180°C (exp)	141°C (exp)
127,3°C (prev)	164°C (prev)	131°C (prev)
<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Liebigs Ann. Chem.</i> , 1984, 108	<i>Liebigs Ann. Chem.</i> , 1984, 108
		
133°C (exp)	64°C (exp)	
169°C (prev)	90,4°C (prev)	
<i>Chem. Ber.</i> , 1971, 104 , 2158	<i>Chem. Ber.</i> , 1971, 104 , 2158	

II-8 - Bibliografia:

- (1) Katritzky, A. R.; Jain, R.; Lomaka, A.; Petrukhin, R.; Karelson, M.; Visser, A. E.; Rogers, R. D.; *J. Chem. Inf. Comput. Sci.*, **2002**, 42, 225-231.
- (2) Zupan, J.; Gasteiger J., *Neural Networks in Chemistry and Drug Design*, 2nd Ed., Wiley-VCH, Weinheim, **1999**.
- (3) Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J. C.; Sheridan, R. P.; Feuston, B. P.; *J. Chem. Inf. Comput. Sci.*, **2003**, 43, 1947-1958.
- (4) Smola, A. J.; Scholkopf, B.; *Statistics and Computing*, **2004**, 14, 199-222.
- (5) Xu, Y.; Zomer, S.; Brereton, R. G.; *Crit. Rev. Anal. Chem.*, **2006**, 36, 177-188.
- (6) Daszykowski, M.; Walczak, B.; Xu, Q. S.; Daeyaert, F.; de Jonge, M. R.; Heeres, J.; Koymans, L. M. H.; Lewi, P. J.; Vinkers, H. M.; Janssen, P. A.; Massart, D. L.; *J. Chem. Inf. Comput. Sci.*, **2004**, 44, 716-726.
- (7) Walden, P.; *Bull. Acad. Imper. Sci. (São Petersburgo)*, **1914**, 1800, citado em: Wasserscheid, P.; Keim, W.; *Angew. Chem. Int. Ed.*, **2000**, 39, 3772-3789.
- (8) Hurley, F. H.; *Electrodeposition of Aluminum*, **1948**, Patente Norte Americana. 2.446.331.
- (9) Chum, H. L.; Kock, V. R.; Miller L. L.; Osteryoung, R. A.; *J. Am. Chem. Soc.*, **1979**, 101, 3264-3265.
- (10) Wilkes, J. S.; Levinsky, J. A.; Wilson, R. A.; Hussey, C. A.; *Inorg. Chem.*, **1982**, 21, 1263-1264.
- (11) Scheffler, T. B.; Hussey, C. L.; Seddon, K. R.; Kear, C. M.; Armitage, P. D.; *Inorg. Chem.*, **1983**, 22, 2099-2100.
- (12) Boon, J. A.; Levinsky, J. A.; Pflug, J. L.; Wilkes, J. S.; *J. Org. Chem.*, **1986**, 51, 480-483.
- (13) Chauvin, Y.; Gilbert, B.; Guibard, I.; *J. Chem. Soc., Chem. Commun.*, **1990**, 1715-1716.
- (14) Wilkes, J. S.; Zaworotko, M. J.; *J. Chem. Soc., Chem. Commun.*, **1992**, 965-967.
- (15) Sheldon, R.; *Chem. Commun.*, **2001**, 2399-2407.
- (16) Freemantle, M.; *Chem. Eng. News*, **1998**, 76 (30th March), 32-37. citado em Earle, M. J.; Seddon, K. R.; *Pure Appl. Chem.*, **2000**, 72, 1391-1398.
- (17) Jain, N.; Kumar, A.; Chauhan, S.; Chauhan, S. M. S.; *Tetrahedron*, **2005**, 61, 1015-1060.
- (18) Chiappe, C.; Pieraccini, D.; *Green Chem.*, **2003**, 5, 193-197.
- (19) Lee, C. K.; Huang, H. W.; Lin, I. J. B.; *Chem. Commun.*, **2000**, 1911-1912.
- (20) Binnemans, K.; *Chem. Rev.*, **2005**, 105, 4148-4204.
- (21) Gordon, C. M.; Holbrey, J. L.; Kennedy, A. R.; Seddon, K. R.; *J. Mater. Chem.*, **1998**, 8, 2627-2636.
- (22) Holbrey, J. D.; Seddon, K. R.; *J. Chem. Soc., Dalton Trans*, **1999**, 2133-2139.
- (23) Godinho, M. H.; Cruz, C.; Teixeira, P. I. C.; Ferreira, A. J.; Costa C.; Kulkarni, P. S.; Afonso, C. A. M.; *Liquid Crystals*, **2008**, 35, 103-107.
- (24) Mathevet, F.; Masson, P.; Nicoud, J-F.; Skoulous, A.; *J. Am. Chem. Soc.*, **2005**, 127, 9053-9061.
- (25) Martin, S. M.; Yonezawa, J.; Horner, M. J.; Macosko, C. W.; Ward, M. D.; *Chem. Matter.*, **2004**, 16, 3045-3055.
- (26) Zhao, H.; *Chem. Eng. Comm.*, **2006**, 193, 1660-1677.
- (27) Antonietti, M.; Kuang, D.; Smarsly, B.; Zhou, Y.; *Angew. Chem. Int. Ed.*, **2004**, 43, 4988-4992.
- (28) Stracke, M. P.; Ebeling, G.; Cataluña R.; Dupont, J.; *Energy Fuels*, **2007**, 21, 1695-1698.
- (29) Zhao, H.; Xia, S.; Ma, P.; *J. Chem. Technol. Biotechnol.*, **2005**, 80, 1089-1096.
- (30) Sheldon, R. A.; Lau, R. M.; Sorgedraeger, M. J.; van Rantwijk, F.; Seddon, K. R.; *Green Chem.*, **2002**, 4, 147-151.
- (31) Freemantle, M.; *Chem. Eng. News*, **2003**, 81(31st March), 9-10.
- (32) Dzyuba, S. V.; Bartsch, R. A.; *J. Heterocyclic Chem.*, **2001**, 38, 265-268.
- (33) Dyson, P. J.; Grossel, M. C.; Srinivasan, N.; Vine, T.; Welton, T.; Williams, D. J.; White, A. J. P.; Zigras, T.; *J. Chem. Soc., Dalton Trans.*, **1997**, 3465-3469.
- (34) Varma, R. S.; Namboodiri, V. V.; *Chem. Commun.*, **2001**, 643-644.
- (35) Dzyuba, S. V.; Bartsch, R. A.; *Chem. Commun.*, **2001**, 1466-1467.
- (36) Schreker, H. S.; Silva, D. O.; Gelesky, G. A.; Stracke, M. P.; Schrekker, C. M. L.; Gonçalves, R. S.; Dupont, J.; *J. Braz. Chem. Soc.*, **2008**, 19, 426-433.

- (37) Wasserscheid, P.; Driessen-Hölscher, B.; van Hal, R.; Steffens, H. C.; Zimmermann, J.; *Chem. Commun.*, **2003**, 2038-2039.
- (38) Zhang, J.; Martin, J. R.; Desmarteau, D. D.; *Chem. Commun.*, **2003**, 2334-2335.
- (39) Blondeau, P.; Segura, M.; Pérez-Fernández, R.; Mendonza, J.; *Chem. Soc. Rev.*, **2007**, 36, 198-210.
- (40) Gao, Y.; Arritt, S. W.; Twamley, B.; Shreeve, J. M.; *Inorg. Chem.*, **2005**, 44, 1704-17112.
- (41) Wang, P.; Zakeeruddin, S. M.; Grätzel, M.; Kantelehner, W.; Mezger, J.; Stoyanov, E. V.; Scherr, O.; *Appl. Phys. A*, **2004**, 79, 73-77.
- (42) Mateus, N. M. M.; Branco, L. C.; Lourenço, N. M. T.; Afonso, C. A. M.; *Green Chem.*, **2003**, 5, 347-352.
- (43) Xie, H.; Zhang, S.; Duan, H.; *Tetrahedron Lett.*, **2004**, 45, 2013-2015.
- (44) Xie, H.; Duan, H.; Li, S.; Zhang, S.; *New J. Chem.*, **2005**, 29, 1199-1203.
- (45) Li, S.; Lin Y.; Xie, H.; Zhang, S.; Xu, J.; *Org. Lett.*, **2006**, 8, 391-394.
- (46) Zhu, A.; Jiang T.; Buxing, H.; Huang, J.; Zhang, J.; Ma, X.; *New J. Chem.*, **2006**, 30, 736-740.
- (47) Xin, X.; Guo, X.; Duan, H.; Lin, Y.; Sun, H.; *Catal. Commun.*, **2007**, 8, 115-117.
- (48) Ishikawa, T.; Kumamoto, T.; *Synthesis*, **2005**, 5, 737-752.
- (49) Branco, L. C.; Serbanovic, A.; da Ponte, M. N.; Afonso, C. A. M.; *Chem. Commun.*, **2005**, 107-109.
- (50) Branco, L. C.; Gois, P. M. P.; Lourenço, N. M. T.; Kurteva, V. B.; Afonso, C. A. M.; *Chem. Commun.*, **2006**, 2371-2372.
- (51) Kantelehner, W.; Haug, E.; Mergen, W. W.; Speh, P.; Maier, T.; Kapassakalidis, J. J.; Bräuner, H.-J.; Hagen, H.; *Liebigs Ann. Chem.*, **1984**, 108-126.
- (52) Kunkel, H.; Mass, G.; *Eur. J. Org. Chem.*, **2007**, 3746-3757.
- (53) Manimala, J. C.; Anslyn, E. V.; *Tetrahedron Lett.*, **2002**, 43, 565-567.
- (54) Lammin, S. G.; Pedgrift, B. L.; Ratcliffe, A. J.; *Tetrahedron Lett.*, **1996**, 37, 6815-6818.
- (55) Yong, Y. F.; Kowalsky, J. A.; Lipton, M. A.; *J. Org. Chem.*, **1997**, 62, 1540-1542.
- (56) Levallet, C.; Lerpiniere, J.; Ko, S. Y.; *Tetrahedron*, **1997**, 53, 5291-5304.
- (57) Dodd, D. S.; Kozikowski, A. P.; *Tetrahedron Lett.*, **1994**, 35, 977-980.
- (58) Schlama, T.; Gouverneur, V.; Valleix, A.; Greiner, A.; Toupet, L.; Mioskowski, C.; *J. Org. Chem.*, **1997**, 62, 4200-4202.
- (59) Seddon, K. R.; Stark, A.; Torres, M. J.; *Pure Appl. Chem.*, **2000**, 72, 2275-2287.
- (60) Nockemann, P.; Binnemans, K.; Driesen, K.; *Chem. Phys. Lett.*, **2005**, 415, 131-136.
- (61) Earle, M. J.; Esperança, J. M. S. S.; Gilea, M. A.; Lopes, J. N. C.; Rebelo, L. P. N.; Magee, J. W.; Seddon, K. R.; Widegren, J. A.; *Nature*, **2006**, 439, 831-834.
- (62) Abramowitz, R.; Yalkowsky, S. H.; *Pharm. Res.*, **1990**, 7, 942-947.
- (63) Silbey, R. J.; Alberty, R. A.; Bawendi, M. G.; *Physical Chemistry*, 4th Ed., John Wiley & Sons, Inc., **2005**.
- (64) Vollhardt, K. P. C.; Schore, N. E.; *Organic Chemistry*, 3rd Ed., W. H. Freeman and Company, New York, **1999**.
- (65) Zhao, L.; Yalkowsky, S.; *Ind. Eng. Chem. Res.*, **1999**, 38, 3581-3584.
- (66) Dannenfelser, R. M.; Yalkowsky, S. H.; *Ind. Eng. Chem. Res.*, **1996**, 35, 1483-1486.
- (67) Karthikeyan, M.; Glen, R. C.; Bender, A.; *J. Chem. Inf. Model.*, **2005**, 45, 581-590.
- (68) Anderson, J. L.; Ding, J.; Welton, T.; Armstrong, D. W.; *J. Am. Chem. Soc.*, **2002**, 124, 14247-14254.
- (69) Eike, D. M.; Brennecke, J. F.; Maginn, E. J.; *Green Chem.*, **2003**, 5, 323-328.
- (70) Holbrey, J. D.; Seddon, K. R.; *J. Chem. Soc., Dalton Trans.*, **1999**, 2133-2139.
- (71) Trohalaky, S.; Pachter, R.; Drake, G. W.; Hawkins, T.; *Energy Fuels*, **2005**, 19, 279-284.
- (72) Santos, L. M. N. B. F.; Lopes, J. N. C.; Coutinho, J. A. P.; Esperança, J. M. S. S.; Gomes, L. R.; Marrucho, I. M.; Rebelo, L. P. N.; *J. Am. Chem. Soc.*, **2007**, 129, 284-285.
- (73) Zhang, S.; Sun, N.; He, X.; Lu, X.; Zhang, X.; *J. Phys. Chem. Ref. Data*, **2006**, 35(4), 1475-1517
- (74) Carrera, G. V. S. M.; Branco, L. C.; Aires-de Sousa, J.; Afonso, C. A. M.; *Tetrahedron*, **2008**, 64, 2216-2224.
- (75) Carrera, G.; Aires-de-Sousa; J. *Green Chem.*, **2005**, 7, 20.

- (76) Katritzky, A. R.; Lomaka, A.; Petrukhin, R.; Jain, R.; Karelson, M.; Visser, A. E.; Rogers, R. D.; *J. Chem. Inf. Comput. Sci.*, **2002**, 42, 71-74.
- (77) Bini, R.; Chiappe, C.; Duce, C.; Micheli, A.; Solaro, R.; Starita, A.; Tine, M. R.; *Green Chem.*, **2008**, 10, 306-309.
- (78) Eike, D. M.; Brennecke, J. F.; Maginn, E. J.; *Green Chem.*, **2003**, 5, 323-328.
- (79) Varnek, A.; Kireeva, N.; Tetko, I. V.; Baskin, I. I.; Solov'ev, V. P.; *J. Chem. Inf. Model.*, **2007**, 47, 1111-1122.
- (80) Descritores Dragon disponíveis em: <http://www.vcclab.org/lab/edragon/start.html>.
- (81) Sun, N.; He, X.; Dong, K.; Zhang, X.; Lu, X.; He, H.; Zhang S.; *Fluid Phase Equilibria*, **2006**, 246, 137-142.
- (82) López-Martin, I.; Burello E.; Davey, P. N.; Seddon, K. R.; Rothenberg, G.; *ChemPhysChem*, **2007**, 8, 690-695.
- (83) Krossing, I.; Slattery, J. M.; Daguenet, C.; Dyson, P. J.; Oleinikova, A.; Weingartner, H.; *J. Am. Chem. Soc.*, **2006**, 128, 13427-13434.
- (84) Jenkins, H. D. B.; Roobottom, H. K.; Passmore, J.; Glasser, L.; *Inorg. Chem.*, **1999**, 38, 3609-3620.
- (85) Jenkins, H. D. B.; Glasser, L.; *Inorg. Chem.*, **2003**, 42, 8702-8708.
- (86) Deetlefs, M.; Seddon, K. R.; Shara, M.; *Phys. Chem. Chem. Phys.*, **2006**, 8, 642-649.
- (87) <http://www.sigmaaldrich.com>
- (88) Toporov, A. A.; Toporova, A. P.; *J. Mol. Struct. (TEOCHEM)*, **2002**, 637, 1-10.
- (89) Hofmann, D. W. M.; *Acta Cryst., Sect B*, **2002**, B57, 489-493.
- (90) Esperança, J. M. S. S.; Guedes, H. J. R.; Blesic, M.; Rebelo, L. P. N.; *J. Chem. Eng. Data*, **2006**, 51, 237-242.
- (91) Xue, H.; Shreeve, J. M.; *Eur. J. Inorg. Chem.*, **2005**, 2573-2580.
- (92) Ye, C.; Shreeve, J. M.; *J. Org. Chem.*, **2004**, 69, 6511-6513.
- (93) Gao, H.; Ye, C.; Winter, R. W.; Gard, G. L.; Sitzmann, M. E.; Shreeve, J. M.; *Eur. J. Inorg. Chem.*, **2006**, 3221-3226.
- (94) Zhao, D.; Fei, Z.; Scopelliti, R.; Dyson, P. J.; *Inorg. Chem.*, **2004**, 43, 2197-2205.
- (95) Matsumoto, K.; Hagiwara, R.; *J. Fluorine Chem.*, **2005**, 126, 1095-1100.
- (96) Ye, C.; Shreeve, J. N.; *J. Phys. Chem. A*, **2007**, 111, 1456-1461.
- (97) Macleod, D. B.; *Trans. Faraday Soc.*, **1923**, 19, 38-41.
- (98) Palomar, J.; Ferro, V. R.; Torrecilla, J. S.; Rodríguez, F.; *Ind. Eng. Chem. Res.*, **2007**, 46, 6041-6048.
- (99) Sadowski, J.; Gasteiger J.; *Chem. Rev.*, **1993**, 93, 2567-2581.
- (100) Sadowski, J.; Gasteiger, J.; Klebe, G.; *J. Chem. Inf. Comput. Sci.*, **1994**, 34, 1000-1008.
- (101) PETRA pode ser testado na página web: <http://www2.chemie.uni-erlangen.de> e disponibilizado pela Molecular Networks GmbH, <http://www.mol-net.de>
- (102) Gasteiger, J.; A Hierarchy of Structure Representations; vol. 3 em *Handbook of Chemoinformatics*; eds J. Gasteiger; T. Engel; Wiley-VCH; New York; **2003**.
- (103) Wagner, M.; Sadowski, J.; Gasteiger, J.; *J. Am. Chem. Soc.*, **1995**, 117, 7769-7775.
- (104) Cálculo de descritores online em: <http://www.vcclab.org/lab/edragon/start.html>.
- (105) Detalhes sobre descritores DRAGON podem ser consultados na página: http://www.taletе.mi.it/dragon_exp.htm.
- (106) Programa R e documentação disponíveis em <http://www.r-project.org/>.
- (107) Aires-de-Sousa, J.; *Chemom. Intell. Lab. Syst.*, **2002**, 61, 167-173.
- (108) Os applets JATOON estão disponíveis em <http://www.dq.fct.unl.pt/staff/jas/jatoon>.
- (109) Vracko, M.; *Curr. Comput. Aided Drug Des.*, **2005**, 1, 73-78.
- (110) Ferreira, R.; Blesic, M.; Trindade, J.; Marrucho, I.; Lopes, J. N. C.; Rebelo, L. P. N.; *Green Chem.*, **2008**, 10, 918-928.
- (111) Scurto, A. M.; Newton, E.; Weikel, R. R.; Draucker, L.; Hallett, J.; Liotta, C. L.; Leitner, W.; Eckert, C. A.; *Ind. Eng. Chem. Res.*, **2008**, 47, 493-501.
- (112) Kulkarni, P. S.; Branco, L. C.; Crespo, J. G.; Nunes, M. C.; Raymundo, A.; Afonso, C.A. M.; *Chem. Eur. J.*, **2007**, 13, 8478-8488.

- (113) Ferreira, R.; Blesic, M.; Trindade, J.; Marrucho, I.; Lopes, J. N. C.; Rebelo, L. P. N.; *Green Chem.*, **2008**, 10, 918-928.

Capítulo III

Aplicação de Métodos de Aprendizagem Automática e Descritores MOLMAP a Problemas de Reactividade Química

III – 1: Introdução Teórica

III - 1.1: Enquadramento

A previsão da reactividade química a partir da fórmula de estrutura dum composto tem ganho importância em diversos contextos nomeadamente nas indústrias química, farmacêutica e cosmética onde é necessário prever a viabilidade de uma dada reacção, a estabilidade química de um dado composto, ou estimar actividades biológicas e toxicidades que estejam relacionadas com a reactividade.

Um caso de toxicidade associada a reactividade química é a dermatite de contacto alérgica em cujo mecanismo existe a formação de ligação covalente entre o composto alérgeno e a cadeia lateral de aminoácido duma proteína. O desenvolvimento de metodologias *in silico* alternativas ao uso de animais em testes de toxicidade é de grande relevância particularmente em resultado da legislação vigente que bane testes em animais para avaliação de produtos cosméticos a partir de 2009 (em casos pontuais até 2013)¹ e da 7ª emenda (2003/15/EC)² à directiva 76/768/CEE da União Europeia relativamente a produtos cosméticos e testes em animais.

O primeiro estudo apresentado neste capítulo consistiu na construção de modelos para a estimativa de reactividade em relação à butilamina. Estes modelos usaram métodos de aprendizagem automática e partiram de uma base de dados de reacções envolvendo a butilamina. A butilamina foi escolhida porque simula a cadeia lateral do aminoácido lisina (que reage nucleofilicamente segundo mecanismo inicial de sensibilização cutânea). Para além da estimativa acerca da possibilidade de reacção ou não, é importante classificar reagentes de acordo com o mecanismo de acção uma vez que são conhecidas relações estrutura-actividade aplicáveis apenas no âmbito de mecanismos específicos. Foram estudados os seguintes aspectos:

- Previsão de reactividade química em relação à butilamina,
- A partir de modelo de reactividade com butilamina, previsões para conjunto de substratos que foram colocados a reagir *in vitro* na presença de lisina³. Desta forma, procurou-se aplicar um modelo obtido a partir da informação extensa de reacções químicas de butilamina, à previsão de reactividade com a lisina, usada para determinar o potencial de um dado composto em provocar dermatite de contacto alérgica.

- Classificação de reacções (na presença de butilamina) segundo mecanismo reaccional.
- Classificação de reagentes (na presença de butilamina) segundo mecanismo reaccional.

Os estudos contemplados neste capítulo usam descritores MOLMAP. Estes foram anteriormente desenvolvidos no laboratório onde este trabalho foi realizado e baseiam-se no mapeamento em rede de *Kohonen* das diversas ligações presentes numa molécula segundo propriedades físico-químicas ou topológicas. Esta metodologia pode ser usada na classificação de reacções, classificação de reagentes e usada na previsão de reactividade. Este tipo de metodologia não necessita da identificação prévia dos centros reaccionais, como é o caso da maior parte dos sistemas de classificação de reacções; nos MOLMAP o centro reaccional pode aparecer implicitamente representado pela diferença entre os MOLMAP de produtos e reagentes. Descritores MOLMAP foram usados com sucesso na classificação de reagentes, produtos e reacções fotoquímicas,⁴ classificação de reacções metabólicas,⁵ na previsão de actividade antioxidante⁶ e previsão de mutagenicidade.⁷

O segundo estudo apresentado neste capítulo consiste na previsão de reactividade química a partir de base de dados de reacções envolvendo cianoborohidreto de sódio (NaCNBH_3). Este é um reagente redutor selectivo amplamente usado em Química Orgânica e foi aqui escolhido para avaliar a capacidade de descritores MOLMAP para diferenciar grupos funcionais semelhantes.

III - 1.2: Classificação de Reacções Químicas

Dada a elevada importância para um químico de um bom sistema de classificação de reacções, tal como ficou patente no tópico anterior, diversos sistemas de classificação têm sido adoptados ao longo dos anos, com complexidade, capacidade de generalização, forma de construção das regras de classificação e propriedades utilizadas na construção das regras diferenciais. Pela forma de obtenção das regras de classificação podem-se distinguir dois tipos fundamentais, por um lado existem os sistemas baseados em modelos rígidos (*Model driven*) segundo regras hierárquicas baseadas na definição de centro reaccional e em alguns casos adicionando informação de átomos e ligações adjacentes a esse centro reaccional. O segundo tipo de sistema, mais recente, consiste na classificação segundo exemplos de reacções específicas (*Data driven*) apresentadas ao sistema. Neste caso a informação é retirada directamente de uma dada base de reacções sem haver uma racionalização prévia e preconcebida de regras/situações na classificação de reacções. Neste último

caso, dependendo das características utilizadas, topológicas e ou físico-químicas, definição explícita ou implícita do centro reaccional, vai ficar definida a flexibilidade e a generalização do sistema. Neste capítulo, para classificação de reacções, são empregues descritores MOLMAP, apresentados mais à frente neste capítulo. Os descritores MOLMAP são usados para a codificação de reagentes e reacções com vista à sua classificação por sistemas *Data driven*. Na construção dos MOLMAP são usadas propriedades físico-químicas e ou topológicas das ligações químicas, não estando definido o centro reaccional, pelo menos de forma explícita, o que permite a aplicação em bases de dados de reacções em que o centro reaccional não está atribuído.

III-1.3: Estado da Arte sobre Métodos Químio-Informáticos para a Previsão de Reactividade Química

Sendo o estudo teórico computacional da reactividade um tópico amplamente estudado, incidiremos aqui em métodos químio-informáticos que se baseiam na aprendizagem automática a partir de conjuntos de dados experimentais e que utilizam métodos da estatística, da neurocomputação ou da inteligência artificial.

Têm surgido diversos métodos/programas que visam prever reactividade, ora reconhecendo estruturas e/ou ligações reactivas, ora estimando possíveis produtos para grau variável de especificação das condições reaccionais, tendo em linha de conta a viabilidade de determinada reacção e/ou a estabilidade do produto, podendo em alguns casos quantificar-se a velocidade, rendimento e proporção de determinado produto.

Podem considerar-se diversos domínios de aplicação, sendo destacados três nesta introdução: previsão de reactividade em reacções orgânicas, geração de metabolitos em sistemas biológicos, previsão de toxicidade e de outras actividades biológicas.

III-1.3.1: Previsão de Reactividade em Reacções Orgânicas

Um dos primeiros métodos de previsão foi o EROS, criado por Gasteiger⁸ sendo referenciado no ano de 1978. Este está ligado a um método de síntese assistida por computador que, para além de construir uma árvore retrosintética até serem atingidos materiais de partida comerciais, restringe as possibilidades de reacção já que cada reacção gerada é avaliada com base em propriedades físico-químicas associadas às ligações relevantes na reacção. Este processo permite reconhecer reagentes pouco reactivos pela via pretendida e produtos pouco estáveis.

Um pouco mais tarde, em 1980, Salatin e Jorgensen concebem o programa de previsão de reactividade CAMEO.⁸ Este programa utiliza conhecimentos mecanísticos para prever o desenrolar da reacção e até mesmo contabilizar a formação de produtos secundários. Centra-se nos conceitos de pKa, nucleofilicidade e electrofilicidade para determinar, por exemplo, se determinada base pode desprotonar determinado composto, ou se pode ocorrer uma eliminação, ou uma reacção de substituição. Nesse processo são também contabilizados factores estéricos. Os produtos gerados neste processo vão ter a sua estabilidade avaliada, constituindo uma medida adicional de viabilidade reaccional.

Em 1993, Simon *et al*⁹ olham para a reactividade química no seu aspecto mais fundamental, centrando-se nas ligações químicas. Os autores consideram nesse estudo ligações reactivas e não reactivas e calculam para cada uma delas 7 propriedades físico-químicas: diferença da carga total entre os dois átomos, diferença da carga π entre os dois átomos, diferença da electronegatividade sigma entre os dois átomos, polaridade da ligação, estabilização por ressonância da quebra heterolítica, polarizabilidade da ligação e energia de dissociação da ligação. É usada em primeira análise uma rede de *Kohonen* de forma a escolher um conjunto de ligações, que cubra um espaço abrangente de forma a ter um conjunto de treino completo e equilibrado. O conjunto de treino assim obtido foi posteriormente usado para treinar uma rede neuronal de retropropagação, método supervisionado que permite prever a reactividade de uma determinada ligação a partir das 7 propriedades físico-químicas.

Em 1996 Satoh *et al*¹⁰ desenvolveram um programa (SOPHIA) de geração de produtos a partir de reagentes e condições reaccionais, estas últimas opcionais. Numa primeira fase é identificado o centro reaccional, sendo posteriormente quebradas as ligações químicas desse mesmo centro. Essas ligações são reconectadas tendo em conta as condições reaccionais e os grupos reactivos e a viabilidades dos produtos gerados é avaliada através de uma base de dados de reacções químicas. Nesta geração de produtos não é usado o conceito de tipo ou categoria de reacção química.

Socorro *et al*¹¹ produzem em 2005 um programa (ROBIA) que a partir de reagentes e condições reaccionais gera produtos reaccionais. Tal como o programa SOPHIA, o programa ROBIA primeiramente identifica o centro reaccional e ligações reactivas e numa segunda fase os produtos reaccionais são gerados a partir de biblioteca de reacções químicas. Posteriormente os produtos gerados são filtrados de forma a eliminar produtos repetidos ou improváveis e finalmente são usados cálculos quânticos para calcular energias dos diversos produtos podendo assim seleccionar os mais estáveis.

III - 1.3.2: Previsão de Reactividade em Reacções Metabólicas

A previsão dos metabolitos dum composto num sistema biológico é uma ajuda importante para estimar toxicidades, bio-activação, persistência, biodegradabilidade ou viabilidade de processos biotecnológicos.

O programa TIMES¹² (*Tissue Metabolism Simulator*) que produz uma série de produtos metabólicos (metabolitos) a partir de determinado composto xenobiótico. O programa possui em base de dados 341 tipos de transformações em mamíferos que aplica segundo as funcionalidades e estrutura do composto em estudo. Posteriormente, o programa avalia a toxicidade, $\log P_{ow}$, e outras propriedades físico-químicas dos metabolitos gerados. Seguindo uma estratégia semelhante, o programa META¹² prevê uma série de biotransformações aeróbicas e anaeróbicas, tal como fotodegradações, compreendendo na sua base de dados 1467 tipos de transformações distintas, sendo obtido uma série de metabolitos para os quais é prevista a sua toxicidade e poucos parâmetros físico-químicos, menos do que no caso do programa TIMES. Disponível no mercado está o programa METEOR¹² que determina uma série de metabolitos associados a um determinado composto xenobiótico para tal usa uma base de dados de 347 tipos de biotransformações que usualmente ocorrem em mamíferos. Os metabolitos gerados são avaliados qualitativamente quanto à probabilidade de aparecimento no organismo, sendo também avaliada a sua toxicidade. No mesmo âmbito surgiu o programa MetabolExpert¹² que utilizando uma biblioteca de 179 tipos de reacções, entre as quais biotransforações associadas ao metabolismo animal e das plantas tal como fotodegradações e degradação associada aos solos, gera uma série de produtos metabólicos a partir de um determinado composto, cada um desses produtos é avaliado quanto à sua toxicidade, $\log P_{ow}$, não determinando nenhuma medida qualitativa nem quantitativa relativamente à prevalência desses compostos no organismo. Mais recentemente surgiu a metodologia, SyGMA,¹³ em que, a partir de uma série de regras associadas a determinadas funcionalidades químicas, são usados descritores *fingerprints* de reacções para classificar e agrupar reacções metabólicas e poder assim refinar as regras existentes e definir novas regras de reactividade metabólica baseadas em conhecimentos extraídos directamente de bases de dados de reacções metabólicas, no final para cada composto é obtida uma série de metabolitos com um grau de probabilidade de aparecimento associada. Por sua vez Takuji *et al*¹⁴ partindo de uma base de dados de reacções enzimáticas do Kegg, vai determinar padrão RDM para cada par reaccional presente nessa base de dados. O padrão RDM codifica o centro R reaccional, a zona M de convergência e a zona D de diferença entre reagente e produto. A partir desta base reaccional codificada desta forma, são testados novos compostos, sendo numa primeira fase pesquisados na base de dados reaccional quais os compostos estruturalmente mais semelhantes, numa segunda fase encontrados os compostos mais semelhantes vão ser aplicadas as transformações

associadas aos padrões RDM associados a determinado composto dessa base de dados sendo obtidos novos compostos que são novamente sujeitos ao mesmo procedimento de forma a obter um percurso metabólico para o composto a ser testado. No mesmo âmbito surgiu o trabalho de Boyer *et al*¹⁵ que utiliza método baseado em Fingerprints associados a 33 tipos de átomos Sybyl, usado para codificar cada átomo do centro reaccional (tendo em conta a sua vizinhança) de cada composto presente numa base de dados metabólica. Tendo em conta essa codificação novos compostos são testados sendo codificados através de Fingerprints de cada átomo particular, cada Fingerprint assim obtido é comparado com os Fingerprints da base de dados metabólica de forma a encontrar Fingerprints equivalentes ao átomo do composto a ser testado. No final para cada átomo do composto em análise é obtido um número de ocorrências, através do programa SPORCalc, correspondendo ao número de vezes que aquele tipo de átomo reagiu na base de dados. Assim, são determinados os átomos do composto em estudo com maior probabilidade em reagir.

III - 1.3.3: Previsão de Toxicidade e outras Actividades Biológicas

Outra área de aplicação relativamente à da reactividade química centra-se na previsão de tipos de toxicidade influenciados pela ocorrência de reacções químicas.

Uma destas propriedades é a mutagenicidade, visto que um dos mecanismos envolvidos consiste na formação de ligações covalentes entre o agente mutagénico e uma base do ADN.¹⁶

Num trabalho produzido no laboratório onde este trabalho foi realizado foram utilizados descritores MOLMAP na construção de modelo *Random Forest* para prever a mutagenicidade (teste de Ames),⁷ modelo que demonstrou boa capacidade de previsão para novos compostos, tendo sido obtida uma percentagem de acertos de 85% para um conjunto de teste de 472 compostos. No mesmo âmbito surgiu recentemente uma metodologia¹⁷ que visa prever o potencial de um composto em provocar aberrações cromossómicas acoplado informação sobre influência directa e indirecta de um composto químico no ADN. O método de previsão baseia-se em dois modelos em série, um que prevê se um composto vai reagir directamente com o ADN. O segundo modelo prevê se o composto vai interagir com topoisomerases (encarregues da replicação do ADN) influenciando assim de forma indirecta as propriedades do ADN.

Também o programa TIMES (descrito anteriormente) após gerar diversos metabolitos a partir do composto em estudo, estima para cada metabolito o potencial para provocar aberrações cromossómicas. Relativamente aos resultados usando esta metodologia, foi obtida uma percentagem de acertos de 80% para compostos que não tinham sofrido activação metabólica. Relativamente a compostos que sofreram activação metabólica foi obtida uma percentagem de acertos de 59%.

Outro aspecto relevante no que diz respeito à toxicidade está relacionado com a reactividade inespecífica de determinados compostos na presença de proteínas contendo grupos tióis. Para estudar o pontencial de um composto em produzir reacções inespecíficas, foi usado um teste *in silico* associado a um teste *in vitro* complementando-se para atingir o mesmo fim.¹⁸ No teste *in silico* foi utilizado um modelo de regressão multilinear para descritores de fragmentos moleculares. O teste *in vitro* consistia no estudo das alterações em espectros de RMN-HSQC H^1/C^{13} e H^1/N^{15} de proteína antigénio La na presença de determinado composto químico (esta é uma proteína estabilizadora de transcritos de ARN protegendo-os de digestão exonucleolítica). Este teste denomina-se ALARM RMN e aproveita a sensibilidade exibida por esta proteína para reagir através de uma cisteína com compostos electrófilos. Este teste *in vitro* foi posteriormente validado através da correlação entre a reactividade exibida pelo teste ALARM RMN e a reactividade exibida por proteínas associadas a toxicidade por reagirem com determinados fármacos, nomeadamente *aldeído desidrogenase*, *superoxidodismutase* e *citocromo P450*, verificando-se uma correlação elevada.¹⁹ Utilizando os resultados do teste ALARM RMN para uma série de 8800 compostos foi construído novo modelo computacional²⁰ usando método de classificação Bayesiano associado a descritores *fingerprints*. Este modelo consegue reconhecer sub-estruturas relevantes para a reactividade na presença do antigene La. O modelo permite obter uma percentagem de acertos de 82% para conjunto de treino de 8800 compostos.

Um exemplo importante de actividade biológica relacionada com reactividade química é a actividade anti-oxidante.²¹ A metodologia MOLMAP foi utilizada para a obtenção de descritores moleculares de compostos fenólicos e construção de dois modelos de actividade anti-radicalar,⁶ um baseado no algoritmo das *Random Forest* para modelo de classificação, o outro baseado em rede neuronal de *Counter-Propagation* para obtenção de modelo de regressão. Foi obtida uma percentagem de 70% de acertos pelo método de validação *leave-one-out* (loo) para modelo de classificação, e uma correlação q^2 de 0,712 pelo método loo para modelo de regressão.

Dermatite de contacto alérgica. A dermatite de contacto alérgica é uma inflamação da pele que resulta da resposta imunitária ao contacto directo de determinados compostos xenobióticos. Até agora as únicas estratégias validadas para identificar de forma conclusiva este tipo de compostos são testes *in vivo* recorrendo ao sacrifício de animais. Por outro lado há a necessidade de substituir testes em animais por outras estratégias alternativas à luz da legislação vigente, tal como a 7ª emenda (2003/15/EC)² à directiva 76/768/CEE da União Europeia. Em conformidade com a 7ª emenda (2003/15/EC), foi calendarizada a supressão gradual de ensaios em animais para avaliação de produtos cosméticos, podendo estes ser realizados até à data de 11 de Março de 2009. Relativamente à colocação

no mercado de produtos cosméticos testados em animais, isso só será possível até 11 de Março de 2013.

Importa conhecer detalhadamente o mecanismo de aparecimento da dermatite de contacto alérgica, para poder desenvolver novos métodos que simulem as condições fisiológicas do organismo humano e que para isso não recorram à experimentação animal. Podem distinguir-se duas fases principais²² (**Figura III.1**). A primeira fase é a da sensibilização, em que um determinado composto alergénio, normalmente electrófilo de baixo peso molecular, entra em contacto com a pele e, após ultrapassar a barreira mais externa da pele (*Stratum corneum*), reage com proteínas/péptidos presentes na pele, nomeadamente glutatona e outras proteínas que contenham aminoácidos com cadeias laterais nucleófilas como é o caso da lisina, histidina e cisteína. Dessa reacção resulta a formação de ligação covalente entre proteína/péptido e composto alergénio. As proteínas/péptidos assim modificadas vão ser identificadas e assimiladas por células de Langerhans, ficando estas células com grupo antigénio incorporado à sua superfície de forma a serem reconhecidas por células T. As células de Langerhans assim modificadas migram desde a epiderme até aos nódulos linfáticos, local onde são reconhecidas pelas células T naïve. A partir desse momento é gerada uma bateria de células T específicas para o composto alergénio em causa, algumas das células T memória são capazes de reconhecer o composto alergénio se este entrar novamente em contacto com a pele.

Há ainda que ter em conta que nem todos os compostos alergénios são electrófilos tendo alguns que ser primeiramente metabolizados/activados no organismo. Posteriormente se a pele for novamente exposta ao agente alérgico, vai ser desenvolvida uma reacção alérgica, reacção essa que passa pelo reconhecimento do composto alérgico por parte de células T memória que entretanto migraram para a região cutânea provenientes dos nódulos linfáticos. Estas células libertam posteriormente um mediador citotóxico que provoca a inflamação do tecido característica da dermatite de contacto alérgica.

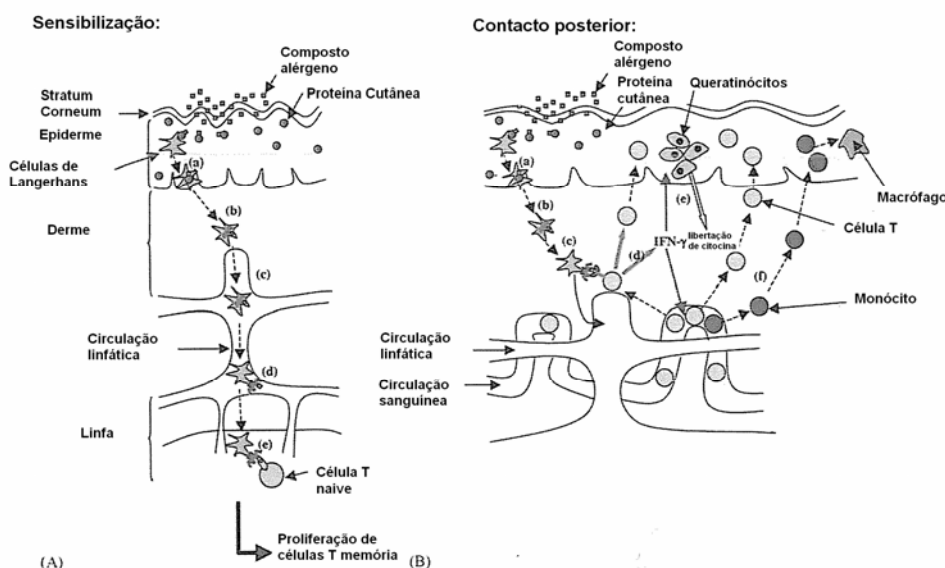


Figura III.1: Representação esquemática das duas fases associadas ao aparecimento da dermatite de contacto alérgica: Sensibilização e Contacto posterior.²²

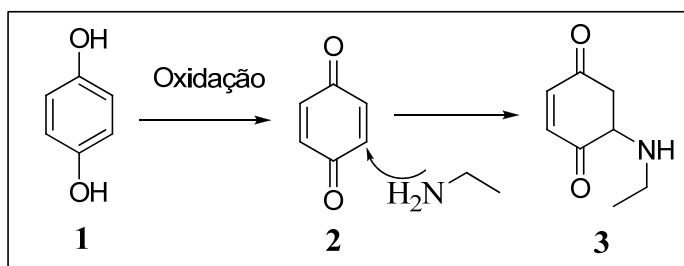
Relativamente a metodologias que recorrem à experimentação animal simulando o comportamento do organismo humano no aparecimento da dermatite de contacto alérgica, existem dois métodos bastante difundidos quer seja pelo seu historial quer seja pela sua actual utilidade. Nesta sequência o método com um historial bastante rico mas que actualmente tem caído em desuso é o teste de sensibilização usando como cobaias porcos da Guiné,²³ que eram expostos através de injeção intradérmica ao composto a avaliar. Este método tem sido substituído por outro teste *in vivo* que reproduz mais fielmente as condições reais em que ocorre a sensibilização cutânea – teste LLNA.²⁴ Este baseia-se na quantificação da resposta ocorrida nos nódulos linfáticos após aplicação tópica de um determinado composto em orelhas de ratos durante 3 dias consecutivos. A quantificação da resposta linfática ocorre ao quinto dia após o início da exposição ao composto em estudo. Nessa altura é introduzida [³H] timidina no organismo do rato através de injeção intravenosa e os ratos são sacrificados sendo excisados os nódulos linfáticos a partir dos quais é medido o grau de incorporação isotópica. O teste estuda directamente a resposta imunitária do organismo ao mesmo tempo que simula as condições reais de exposição tópica e penetração do composto na pele, algo que não acontecia com o teste realizado em porcos da Guiné. Este teste foi adoptado recentemente pela OCDE²⁵ como método utilizado para avaliar o potencial de um composto químico em provocar sensibilização cutânea.

No sentido de provocar ruptura com o *status quo*, em que o teste animal é a forma mais fiável e implementada, foram surgindo novos testes mais ou menos gerais, *in vitro*, *in silico* e até mesmo racionalizações estruturais e mecanísticas, usados separadamente ou em conjugação.

Para racionalizar a relação entre a estrutura química/mecanismo reaccional e potencial para reagir com a pele, Aptula *et al*¹ considera central, no aparecimento da dermatite de contacto alérgica, a reacção de formação de ligação covalente entre composto alérgeno electrófilo e proteína cutânea nucleófila. Estes autores agrupam uma série de 41 compostos em diferentes grupos segundo o mecanismo reaccional associado à reacção com proteínas da pele. Os diferentes grupos estão constituídos por compostos que reagem como aceitadores de Michael, compostos que reagem em substituição nucleofílica aromática (S_NAr), substituição nucleofílica, compostos que vão formar bases de Shiff, agentes acilantes, e finalmente compostos que não exibem características electrófilicas evidentes. Os autores analisam cada um dos grupos individualmente, como detalhado de seguida.

Reacções com aceitadores de Michael:

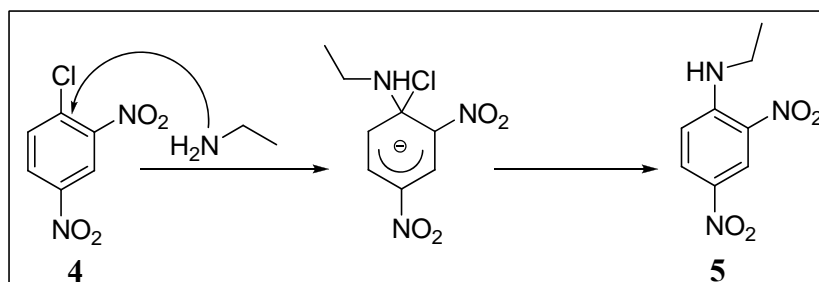
O ataque nucleofílico dá-se na ligação dupla C-C quando activada por um grupo electroatractor como no caso de **2** no **Esquema III.1**, havendo os casos em que o aceitador de Michael tenha de ser previamente activado (caso da hidroquinona **1**).



Esquema III.1: Activação de aceitadores pro-Michael (Oxidação).

Substituição nucleofílica aromática (S_NAr):

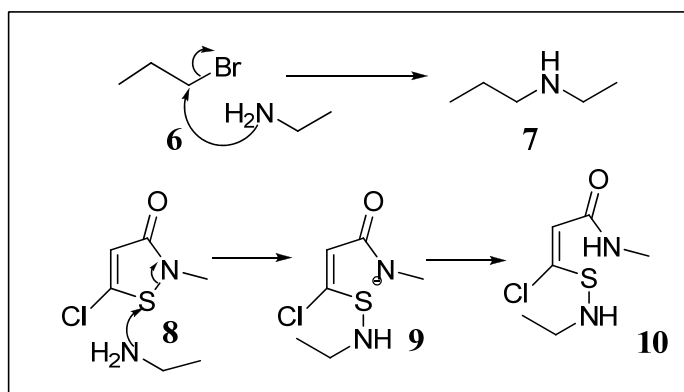
Neste tipo de reacção ocorre ataque nucleófilo num carbono aromático contendo grupo electronegativo que ao mesmo tempo possa ser grupo de saída (como é o caso do cloro em **4**). A carga negativa no intermediário é estabilizada pela ressonância de grupos electroatractores, no caso do **Esquema III.2** grupos nitro. Na última fase dá-se a saída do cloreto e é recuperada a aromaticidade.



Esquema III.2: Reacção $\text{S}_{\text{N}}\text{Ar}$.

Substituição nucleofílica:

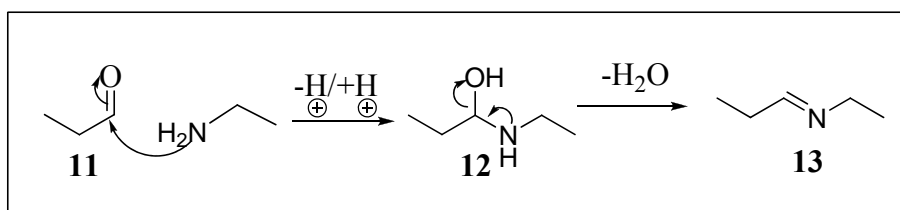
Haloalcanos **6** são exemplos característicos de electrófilos típicos de reacções $\text{S}_{\text{N}}2$ mas existem outros tipos de electrófilos tal como a 5-cloro-2-metilisotiazol-3-ona **8** que pode actuar como sensibilizador da pele.



Esquema III.3: Substituições nucleofílicas.

Electrófilos Capazes de Formar Bases de Schiff:

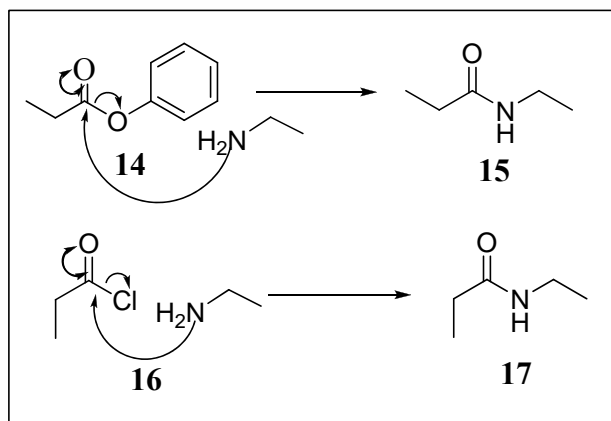
Nesta família de reacções enquadram-se como electrófilos, aldeídos e cetonas activadas sendo mais reactivos os carbonilos electro-deficientes.



Esquema III.4: Formação de base de Schiff 13.

Electrófilos Agentes Acilantes:

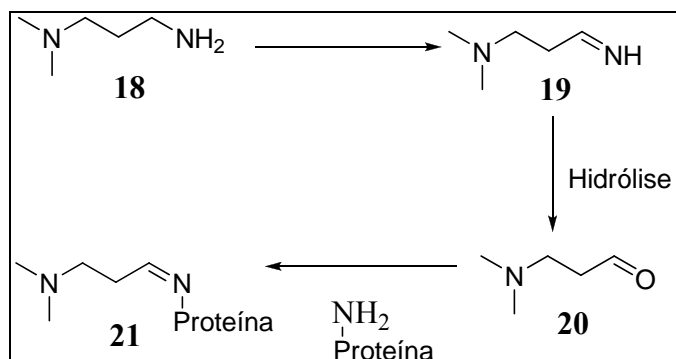
Ésteres de álcoois alifáticos simples não são agentes acilantes pois o correspondente alcóxido é demasiado básico para ser um bom grupo abandonante. Ésteres de álcoois mais ácidos tal como os fenóis (**14**) e anidridos carboxílicos são electrófilos e podem acilar nucleófilos. Haletos de acilo **16** são também bons agentes acilantes.



Esquema III.5: Electrófilos actuando como agentes acilantes.

Compostos que não Exibem Características Electrófilas Evidentes:

Este é um grupo estruturalmente muito diverso, em que a maior parte dos compostos são fracos sensibilizadores cutâneos ou não o são de todo. Aqueles que apresentam alguma reactividade normalmente são previamente activados metabolicamente como é o caso da 3-dimetilaminopropilamina:



Esquema III.6: Activação metabólica da 3-dimetilaminopropilamina **18** e posterior reacção com proteína cutânea.

Mais recentemente²⁶ os mesmos autores aplicaram a mesma estratégia a um conjunto mais abrangente de 106 compostos que haviam sido previamente testados no teste LLNA. Esse mesmo conjunto já tinha sido anteriormente analisado por Ashby *et al*²⁷ mas de um ponto de vista mais estrutural do que mecanístico.

Um tipo de abordagem diferente consiste na utilização de metodologia *in silico*, que visa prever a partir da estrutura de um dado composto se este vai provocar dermatite de contacto alérgica. É um conceito bastante atractivo não só por poupar vidas a animais, mas também pela rapidez e baixo custo associados à obtenção de resultados. Uma das primeiras propostas neste sentido foi o modelo RAI²⁶ em 1982, que se baseia numa fórmula simples, função da dose do composto alérgeno D , constante cinética da reacção entre o composto sensibilizador e as proteínas da pele k e hidrofobicidade do composto P , sendo P o coeficiente de partição entre octanol e água:

$$RAI = \log D + a \log k + b \log P$$

Equação III.1

Este modelo foi concebido para simular o teste *in vivo* recorrendo a porcos da Guiné.

Entretanto muitos outros modelos *in silico* foram desenvolvidos, alguns deles comerciais como é o caso dos programas DEREK, TOPKAT, MCASE, TIMES-SS e BIOASAR.²⁸ Alguns destes modelos baseiam-se em regras estruturais/mecanísticas simples, como é o caso dos programas DEREK e TIMES-SS, ou também têm incorporado na sua base de dados reacções de activação metabólica como é o caso do programa TIMES-SS, mas estão ainda longe de atingir a mesma capacidade de avaliação de compostos alérgenos do teste *in vivo* LLNA.

Ainda descritos na literatura aparecem modelos específicos para determinados tipos de famílias particulares em termos mecanísticos e estruturais, como é o caso de um modelo descrito por Roberts,²⁹ centrado em compostos aromáticos que reagem na presença de composto nucleofílico segundo mecanismo de substituição nucleofílica aromática (S_NAr). Este modelo usa os parâmetros de Hammet e Taft associados a grupos substituintes em posições relativas, *orto*, *meta*, *para* e *ipso*. A equação aparece no seguinte formato:

$$RP = \sum \sigma^- (\text{orto}, \text{meta}, \text{para}) + 0.45 \sigma^* (\text{ipso})$$

Equação III.2

Esta equação contabiliza ainda se o composto é heteroaromático através do σ^- associado à posição relativa em que está situado o heteroátomo. Os resultados do modelo correlacionam bem com dados experimentais obtidos no teste *in vivo* utilizando porcos da Guiné.

Patlewicz *et al*³⁰ propuseram dois modelos de reactividade distintos, um dos modelos é dirigido a compostos aceitadores de Michael e o outro relativo a compostos que formam bases de Schiff. Ambos os modelos são regressões multi-lineares (MLR) contabilizando factores de hidrofobicidade e grupos substituintes – no caso de aldeídos aceitadores de Michael grupos α e β em relação a carbonilo. Ambos os modelos apresentam problemas para compostos com conjugação adicional, obtendo-se na generalidade previsões acertadas. As regras obtidas para estes dois modelos foram incorporadas no programa comercial DEREK.

Mais recentemente foi construído um modelo geral *Random Forest*,³¹ obtido a partir de base de dados de 178 compostos orgânicos aos quais tinha sido aplicado o teste *in vivo* LLNA. Os compostos foram classificados como reactivos ou não reactivos. A metodologia utilizada é de algum modo surpreendente já que apesar de as *Random Forest* lidarem bem com grandes conjuntos de descritores, os autores demonstram que é possível melhorar a capacidade de previsão do modelo através de uma escolha judiciosa dos descritores mais eficazes, *i. e.*, nesse método os autores partiram de um conjunto de 1380 descritores moleculares e construíram um modelo *Random Forest* do qual foi obtida uma medida dos descritores mais relevantes, sendo assim seleccionados metade dos descritores iniciais, que foram utilizados para construir novo modelo *Random Forest*. Este procedimento foi repetido sucessivas vezes na construção de modelos *Random Forest* que tivessem como base um conjunto de descritores progressivamente mais reduzido, sendo assim encontrada a *Random Forest* com melhor capacidade de previsão. De forma a melhorar ainda mais essa capacidade, utilizou-se o conjunto de descritores usados para construir a *Random Forest* imediatamente anterior à melhor para ir descartando um a um os descritores menos importantes para o modelo, obtendo-se assim uma sucessão de modelos *Random Forest* até encontrar o modelo óptimo que permitiu a obtenção de uma percentagem de erros de 16,3% numa espécie de validação cruzada (*Out of Bag* - explicado no Capítulo I no âmbito do algoritmo da *Random Forest*). Este modelo está constituído por 10 descritores moleculares. Como explicado no Capítulo I, *Random Forest* permitem a obtenção de uma medida de proximidade entre os compostos, medida essa que foi aqui utilizada em método de *clustering* para agrupar compostos com características semelhantes.

Entre os métodos *in vivo* e os métodos *in silico* foram criados e optimizados métodos *in vitro*. Já em 1936 Landsteiner e Jacobs³² tinham racionalizado o fenómeno de aparecimento da dermatite de contacto alérgica a partir de formação de ligação covalente entre composto alérgeno e proteínas da pele. Para demonstrar esse facto realizaram teste *in vitro* pondo a reagir um nucleófilo

modelo, a anilina, que simulava as proteínas cutâneas, e determinado composto electrófilo, neste caso compostos aromáticos em que ocorria substituição nucleofílica aromática. Posteriormente compararam os resultados deste teste com os resultados obtidos, através do teste *in vivo* utilizando porcos da Guiné como cobaias, observando-se uma correlação perfeita entre os resultados dos dois testes. Mais recentemente, Aptula *et al*³³ descreveram o uso de um conjunto de dois testes *in vitro* para encontrar substâncias químicas alérgicas. Um dos testes consiste no estudo espectrofotométrico de uma solução em que foi previamente posto a reagir o péptido glutatona e a substância electrófila a avaliar. Os grupos tiol livres provenientes da glutatona são avaliados a 412 nm depois de terem sido derivatizados com ácido 5,5'-ditio-bis(2-nitrobenzóico) (DTNB). O outro teste é um ensaio espectrofotométrico para detectar citotoxicidade de determinado composto na presença de *Tetrahymena pyriformis* (TETRATOX) obtida através da quantificação da inibição de crescimento da população deste microorganismo. Na avaliação da toxicidade geral há que ter em conta um tipo de toxicidade não específica não relacionada com ataque nucleófilo de proteínas à substância em estudo, a qual está correlacionada com a hidrofobicidade da molécula, sendo descontada da toxicidade geral evidenciada pelo teste espectrofotométrico. Para determinar esta toxicidade não específica é usada uma equação. No final a combinação dos resultados dos dois testes permite boa correlação com os resultados obtidos pelo teste LLNA.

Mais recentemente Natsch *et al*³⁴ usam informação referente a uma bateria de testes *in vitro*, essa informação é comparada com resultados referentes a teste *in vivo* LLNA sendo obtida uma correlação de 87,9%.

Por sua vez, Gerberick *et al*^{3,35} haviam criado e optimizado uma série de testes *in vitro* colocando a reagir o composto em estudo e um dado péptido contendo um dado aminoácido nucleófilo – lisina, cisteína, histidina, ou glutatona. Nesses estudos, após período pré-definido de reacção entre péptido e composto a analisar, a mistura reaccional foi eluída por HPLC, sendo detectada e quantificada por detector de foto-díodos a fracção de péptido que não reagiu. Foram testados diversos compostos químicos que também foram testados *in vivo* no ensaio LLNA. Comparou-se os resultados obtidos entre o teste *in vivo* LLNA e cada um dos testes *in vitro* e verificou-se³ que os resultados correspondiam em 83,8% dos casos, no caso de o teste *in vitro* se basear na reactividade de péptido contendo cisteína, 66,7% no caso do péptido conter lisina, 65,8% no caso da glutatona e 36,1% no caso de péptido conter histidina como aminoácido nucleófilo.

Baseada neste estudo,³ Almeida³⁶ explorou métodos de aprendizagem automática para prever a reactividade de determinado composto na presença de glutatona, tendo como base um pequeno conjunto de dados extraído de base de dados de reacções com glutatona. O modelo foi validado com

um conjunto de teste de 50 compostos sendo obtida uma percentagem de acertos de 72%. Na obtenção deste modelo foram usados descritores MOLMAP.

Nesta tese foi construído um modelo de previsão de reactividade na presença de butilamina, nucleófilo modelo; uma das aplicações desse modelo visa ter acesso ao potencial de um determinado composto em provocar dermatite de contacto alérgica. Seguiu-se esta abordagem visto um dos aminoácidos presentes na pele ser a lisina, sendo a cadeia lateral deste aminoácido, responsável pela sua nucleofilicidade, semelhante à butilamina (**Figura III.2**).

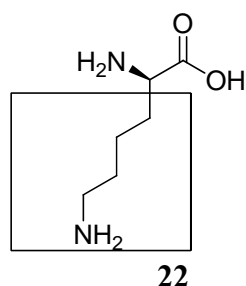


Figura III.2: Aminoácido L-lisina e o pormenor da sua cadeia lateral, semelhante à butilamina.

Por outro lado, os resultados dos ensaios *in vitro* realizados por Gerberick³ na presença de péptido contendo lisina revelaram uma correspondência elevada com resultados de teste *in vivo* LLNA, o que justifica explorar um modelo de reactividade na presença de butilamina.

A previsão de reactividade é nos dias de hoje um tópico bastante relevante nomeadamente na indústria química, farmacêutica ou cosmética onde é necessário prever a viabilidade de uma reacção, estimar toxicidades, actividades biológicas ou estabilidade, nesse sentido foram surgindo diversos métodos/programas que visavam prever reactividade, ora ao nível de reconhecimento de estruturas e/ou ligações reactivas, ora na estimativa de possíveis produtos para grau variável de especificação das condições reaccionais, tendo em linha de conta a viabilidade de determinada reacção e/ou a estabilidade do produto, podendo em alguns casos quantificar-se a velocidade, rendimento e proporção de determinado produto.

No que à reactividade diz respeito podem-se considerar diversos domínios de aplicação, nesta introdução serão destacadas quatro: Previsão de reactividade em reacções orgânicas, geração de metabolitos, previsão de toxicidade e previsão de outras actividades biológicas.

III - 1.4: Descritores MOLMAP

A metodologia MOLMAP (*MO*Lecular *MA*ps of *At*om-level *Pr*operties) visa a obtenção de codificação numérica de estruturas moleculares, e reacções.⁴ Genericamente, o método permite a obtenção do perfil das ligações contidas numa determinada molécula baseando-se no mapeamento, em rede neuronal de *Kohonen*, de cada uma das ligações da molécula, podendo estas estar representadas por propriedades físico-químicas ou topológicas. Por outro lado, este método permite representar reacções sem explicitação do centro reaccional podendo este ser implicitamente representado pela diferença entre os MOLMAPs de produtos e reagentes. Neste trabalho foram usados MOLMAPs de reagentes e também MOLMAP de reacções.

No processo que visa a obtenção de MOLMAP há que considerar duas fases distintas, a primeira consiste numa fase de treino onde é obtido um mapa de ligações química que, na segunda fase, é usado para mapear as ligações químicas de uma determinada molécula, obtendo um perfil de ligações químicas de determinada molécula, ou seja, um MOLMAP.

Para a concretização da primeira fase parte-se de um conjunto de ligações (associada cada ligação a determinado composto), sendo calculadas uma série de propriedades físico-químicas e/ou topológicas para cada ligação, nesses cálculos é tido em linha de conta o ambiente químico de cada ligação. No final vão ser obtidas uma série de ligações químicas, cada uma representada por um vector contendo propriedades dessa ligação, que foram usadas para treinar uma rede neuronal de *Kohonen* (**Figura III.3**) que depois de treinada permite a obtenção de um mapa de ligações químicas. Durante o treino as ligações químicas, representadas cada uma por um vector de propriedades, vão ser submetidas sequencialmente à rede de *Kohonen*, neste caso para uma determinada ligação vai ser encontrado o neurónio vencedor, aquele cujos pesos da camada de *input* são mais semelhantes às propriedades representadas no vector dessa ligação em termos de uma medida directamente relacionada com a distância Euclidiana (mais pormenores sobre a arquitectura e descrição do procedimento de treino e classificação das redes neuronais de *Kohonen* no primeiro capítulo desta tese). Aqui há que ter em conta que cada coluna na rede de *Kohonen* representa um determinado neurónio (**Figura III.3**), estando composto por diversos pesos, tantos quanto o número de propriedades de cada ligação. Posteriormente, os pesos desse neurónio vão ser modificados de forma a ficarem mais próximos das propriedades dessa ligação. Mas não só os pesos do neurónio vencedor vão ser alterados também os pesos dos neurónios vizinhos vão ser alterados, ainda que não de uma forma tão pronunciada. O mesmo procedimento repetiu-se para as restantes ligações escolhidas, sendo estas submetidas repetidas vezes à rede de *Kohonen*, tantas como o número pré-definido de *epochs*. No final vai ser obtido um mapa de *Kohonen* treinado com um aspecto semelhante ao da **Figura III.4** onde podem estar mapeadas uma série de ligações. Um dos sinais para o treino de um mapa *Kohonen*

bem sucedido consiste no mapeamento em zona próxima de ligações semelhantes como se verifica no mapa da **Figura III.4**.

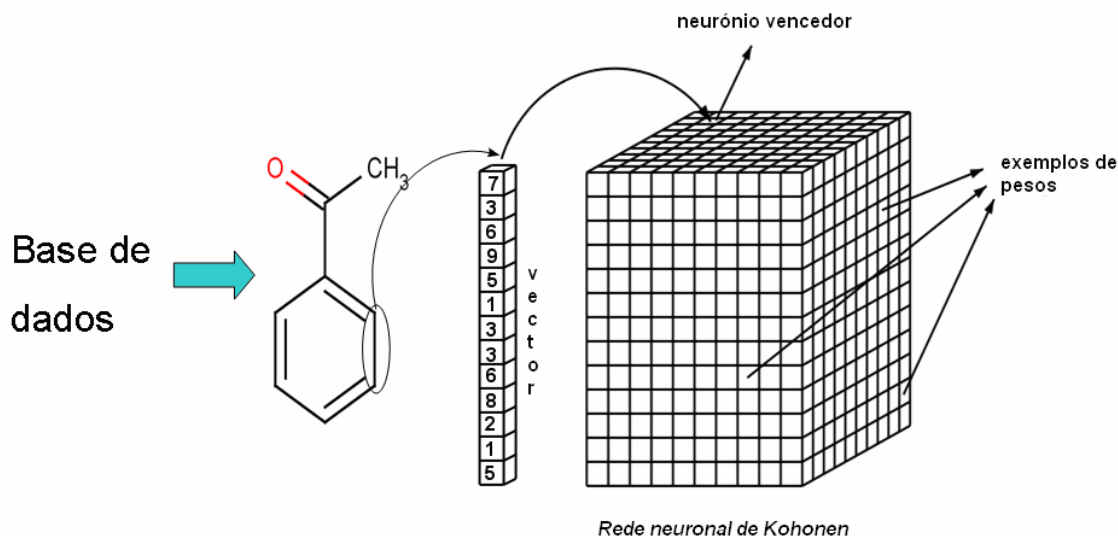


Figura III.3: Exemplo esquemático do processo de treino, que visa a obtenção de um mapa de ligações químicas como o da Figura III.4.

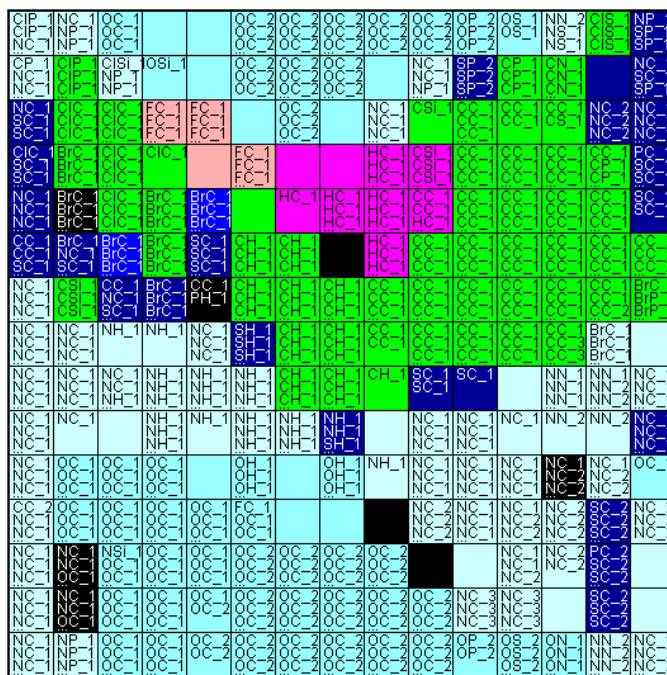


Figura III.4: Exemplo de um mapa de *Kohonen* após a fase do treino com uma série de ligações químicas mapeadas representadas pelos elementos químicos e pela ordem dessa ligação. Por exemplo a ligação CC_1 representa ligação simples carbono-carbono.

Na segunda fase, cada uma das moléculas do conjunto considerado em determinado estudo vai ser submetida ao mapa de *Kohonen* treinado sendo obtido um perfil das ligações dessa molécula, representado cada tipo de ligação presente nessa molécula por um “x” na **Figura III.5**.

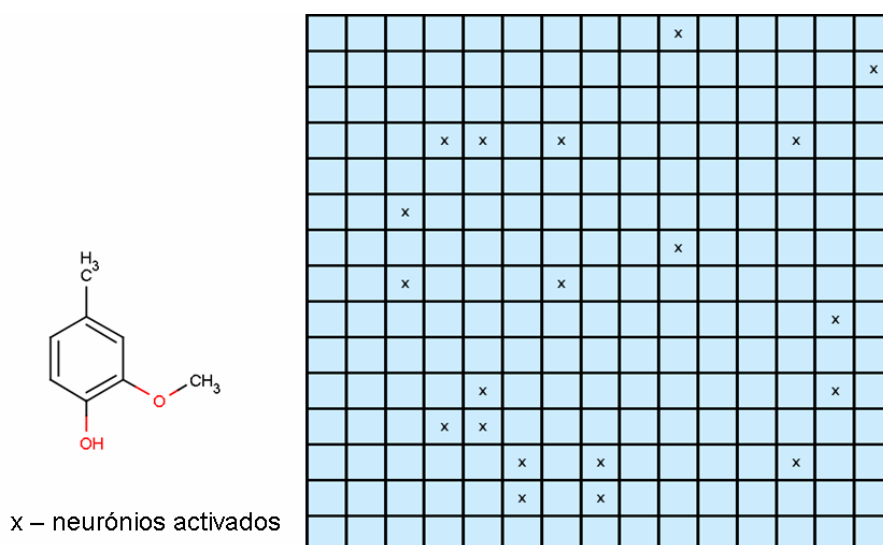


Figura III.5: Perfil das ligações químicas de uma determinada molécula, representado cada tipo de ligação por um “x” no mapa de *Kohonen*.

O perfil das ligações activadas no mapa de *Kohonen* pode ser substituída por uma representação numérica (**Figura III.6**), em que cada número representa a abundância de determinado tipo de ligação nessa molécula, sendo assim obtida uma representação de uma determinada molécula denominando-se MOLMAP. Esse MOLMAP pode ser representado por um vector (parte de baixo da **Figura III.6**) concatenando cada uma das colunas do MOLMAP da esquerda para a direita.

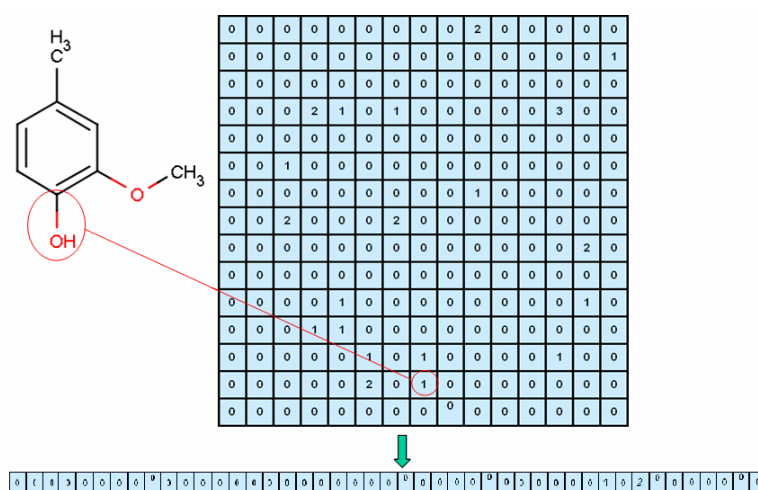


Figura III.6: Representação do perfil das ligações químicas de uma molécula na forma numérica em que cada número representa a abundância de determinado tipo de ligação.

III - 1.5: Derivação de Informação Negativa a Partir de Bases de Dados Reaccionais Compostas Exclusivamente por Reacções que Ocorrem

De forma a treinar um modelo que preveja se um composto reage ou não, são necessários compostos que reagem (positivos) e compostos que não reagem (negativos). No entanto as bases de dados reaccionais normalmente não incluem exemplos de reacções falhadas, *i.e.*, não existe informação explícita sobre compostos que não reagem com um reagente específico. Na falta dessa informação explícita é possível obtê-la através de subestruturas de determinado reagente que não incluam o centro reaccional.³⁷ Na maior parte das reacções, os reagentes têm um único local para ocorrer reacção com um determinado agente (butilamina ou cianoborohidreto de sódio no caso específico do trabalho descrito nesta dissertação). Regiões da molécula não envolvidas na reacção podem ser usadas como exemplos negativos.

A obtenção de informação negativa, baseada na codificação de sub-estruturas de um composto que permanecem inalteradas durante determinada reacção, pode ser concretizada recorrendo à metodologia MOLMAP, ou seja, comparando o MOLMAP de um reagente ao MOLMAP do respectivo produto aparecem evidenciados componentes de MOLMAP (ligações) cujo valor no produto é igual ou superior ao verificado no reagente, estes componentes correspondem a ligações químicas, com propriedades físico-químicas e topológicas praticamente inalteradas e que geralmente podem ser consideradas ligações que não reagem. Um MOLMAP codificando as ligações do reagente que permanecem durante a reacção pode ser usado como MOLMAP de um pseudocomposto que não reage, um exemplo negativo requerido para o treino de um modelo de previsão de reactividade química.

III-2: Metodologia Computacional

No trabalho descrito ao longo deste capítulo são considerados 4 tipos de modelos: **a)** Modelos de classificação de compostos segundo mecanismo reaccional na presença de butilamina, **b)** Modelos de classificação de reacções segundo mecanismo reaccional estando envolvida a butilamina, **c)** Modelos de previsão de reactividade de compostos na presença de butilamina (nucleófilo modelo) e **d)** Modelos de previsão de reactividade de compostos na presença de cianoborohidreto de sódio.

III-2.1: Bases de Dados Reacionais Originais

Na construção dos 4 tipos de modelos usou-se como fontes de informação duas bases de dados de reacções extraídas da base de dados SPRESI (Infochem³⁸ GmbH, Munique, Alemanha). Uma das bases de dados consiste num conjunto de 3671 reacções em que a butilamina intervém como reagente; a segunda base de dados consiste num conjunto de 2571 reacções em que o cianoborohidreto de sódio intervém como reagente. Ambas as bases de dados foram estandardizadas e sujeitas a processo de selecção de reacções.

III-2.1.1: Estandardização de Reacções

Para assegurar uniformidade nos critérios de comparação entre os diferentes objectos em estudo (compostos ou reacções) todos os compostos intervenientes em cada uma das reacções provenientes das 2 bases de dados usadas neste estudo foram estandardizados usando o programa Standardize do ambiente JCHEMA (versão 3.2.2),³⁹ com os seguintes critérios:

- Considerar de forma explícita todos os átomos de hidrogénio presentes em cada composto.
- Para o caso dos estudos **a)** e **b)** as ligações aromáticas foram consideradas como tal visto o programa cxcalc do ambiente MARVINBEANS aceitar ligações aromáticas. No caso dos estudos **c)** e **d)** as ligações aromáticas não são explicitamente marcadas como tal para serem aceites pelo programa Petra (versão 4) para o cálculo de propriedades físico-químicas de ligações químicas.
- Não considerar estereoquímica associada a cada molécula.

III-2.1.2: Selecção de Reacções

Do conjunto de 3671 reacções em que a butilamina intervém como reagente foram seleccionadas reacções que respeitavam as seguintes condições:

- Reacções só com um substrato e um equivalente de butilamina como reagentes.
- Reacções com apenas um produto.
- Não poderá intervir nenhum tipo de catalisador nas reacções consideradas.
- Reacções estequiometricamente equilibradas com uma tolerância de (0 – +6) átomos comparando reagentes em relação ao produto.
- Reacções cujos compostos intervenientes fossem compatíveis com os programas/ambientes Petra (versão 4),⁴⁰ Marvin (versão 4.1.4)³⁹ e JATOON.^{41,42}

Das 3671 reacções iniciais permaneceram 971. Essas reacções foram distribuídas/processadas de forma diferencial pelos diferentes tipos de estudo: **a)**, **b)** e **c)**.

A segunda base de dados utilizada consistia num conjunto de 2571 reacções em que o cianoborohidreto de sódio intervém como reagente; desse conjunto foi seleccionado um grupo de reacções que cumpriam os seguintes critérios:

- Reacções só com um substrato e um equivalente de cianoborohidreto de sódio como reagentes.
- Reacções com apenas um produto.
- Não poderá intervir nenhum tipo de catalisador nas reacções consideradas.
- Reacções estequiometricamente equilibradas com uma tolerância de (+6 – -8) átomos comparando substrato em relação ao produto.
- Reacções cujos compostos intervenientes fossem compatíveis com os programas/ambientes Petra (versão 4) e JATOON.

Cumprindo estes critérios permaneceram 364 reacções que foram usadas no estudo **d)**.

III-2.2: Descritores Moleculares MOLMAPs

A obtenção de descritores MOLMAPs divide-se em duas partes fundamentais, a primeira consiste na obtenção de determinado mapa de ligações químicas baseado no treino de uma rede neuronal de *Kohonen*, a segunda consiste na obtenção do perfil de ligações de determinado composto no mapa anteriormente obtido.

III-2.2.1: Mapeamento de ligações

A partir da base de dados de 971 reacções em que intervém butilamina procedeu-se à separação de reagentes e produtos (a butilamina foi descartada) obtendo-se assim 1942 compostos. Todos os compostos foram submetidos aos programas Petra (versão 4) e *cxcalc* (proveniente do ambiente Marvin versão 4.1.4) para cálculo de propriedades físico-químicas e ou topológicas para cada uma das ligações de cada molécula. Nesse cálculo cada ligação ficou representada duas vezes pois certas propriedades vão ser diferentes conforme a ligação seja representada na forma A-B ou B-A, sendo considerada a representação de ligação que comesçasse com o átomo de maior carga total entre os dois átomos da ligação. Assim, estava-se a seguir o mesmo critério de representação para poder comparar ligações distintas. Entre todas as ligações dos 1942 compostos foram consideradas apenas aquelas em que algum dos átomos da ligação seja heteroátomo ou pertença a ligação múltipla, ou em que algum átomo vizinho a essa ligação seja heteroátomo ou pertença a ligação múltipla.

Entre todas as ligações dos 1942 compostos foram seleccionados de forma aleatória subconjuntos para treinar a rede de Kohonen que cobrissem todo o universo de tipos de ligações presentes na base de dados aqui em causa, e adicionalmente foi tida em consideração a abundância de cada tipo de ligação e a relevância desse tipo de ligação para os estudos de reactividade/determinação de mecanismo reaccional. Cada uma das ligações seleccionadas foi usada como objecto para treinar redes neuronais de *Kohonen* de diferentes dimensões e assim obter mapas de ligações. Foi seleccionado um conjunto de ligações diferente para os diferentes estudos considerados neste capítulo. Tendo em conta os resultados obtidos no mapeamento de 5870 ligações químicas nos estudos a) e b), procurou-se nos estudos c) e d) aumentar o número de carbonilos e equilibrar o número de ligações C-X, em que X é um halogénio, de forma a que os mapas tivessem uma capacidade de discriminação optimizada.

a) estudo do mecanismo reaccional de compostos na presença de butilamina

Foram seleccionadas 5870 ligações químicas.

b) estudo mecanismo reaccional associado a reacções em que intervenha a butilamina

Foi seleccionado o mesmo conjunto de 5870 ligações usadas no estudo a).

c) estudo de reactividade de compostos na presença de butilamina

Foram seleccionadas 4980 ligações químicas.

d) estudo de reactividade de compostos na presença de cianoborohidreto de sódio

Foi seleccionado o mesmo conjunto de 4980 ligações usadas no estudo c).

Cada ligação seleccionada foi representada por diversos subconjuntos de descritores, que foram normalizados através da normalização Z que consiste na aplicação da seguinte expressão a cada um dos descritores usados:

$$Descriptor_normalizado = \frac{(Descriptor - media)}{desvio_padrão}$$

Equação III.3

Neste caso *media* e *desvio padrão* são o valor médio e desvio padrão associado a determinado tipo de descritor considerando o conjunto total de ligações usadas para construir os mapas de ligações. *Descriptor* é o valor correspondente a determinado objecto referente a determinado tipo de descritor. Posteriormente foram testados os diversos subconjuntos de descritores no treino de redes de *Kohonen* para a obtenção de mapas de ligações. O critério para a escolha dos subconjuntos de descritores mais adequados para os estudos em causa baseou-se na comparação manual da capacidade de separação, de diversas ligações semelhantes pertencentes a diferentes funcionalidades/ambientes químicos, evidenciada por cada mapa de ligações treinado com conjunto de descritores específico. Foram comparados mapas 29×29 através do mapeamento de 22 moléculas distintas. Quinze moléculas são compostos carbonílicos estando cobertas funcionalidades diversas (ácidos carboxílicos, amidas,

ésteres, cloretos de ácido, cetonas e aldeídos) e ambientes químicos. Foi avaliada a capacidade para distinguir os diferentes carbonilos. As 15 moléculas são as seguintes:

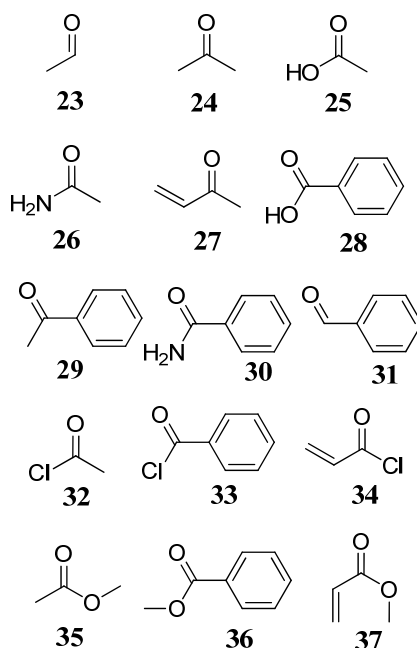


Figura III.7: 15 compostos contendo ligação carbonilo constituindo diferentes funcionalidades e abrangendo diversos ambientes químicos.

As restantes 7 moléculas foram analisadas em separado visto cada uma conter diferentes tipos de ligações. A seguinte molécula a ser analisada foi a seguinte:

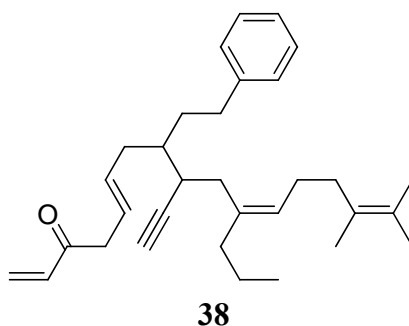


Figura III.8: Molécula tendo em evidência diferentes ligações carbono-carbono, ligações simples, duplas, triplas e aromáticas.

Foram mapeadas nos diferentes mapas em comparação as ligações carbono-carbono presentes nesta molécula sendo avaliada a capacidade para distinguir entre ligações simples, duplas, triplas e aromáticas sendo tido em linha de conta o ambiente químico de cada ligação em estudo. A seguinte molécula a ser analisada foi a seguinte:

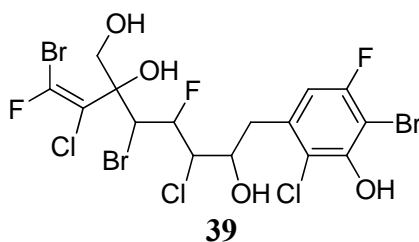


Figura III.9: Molécula onde estão evidenciadas diversas ligações carbono-halogénio e carbono-oxigénio.

Relativamente à estrutura **39** foram mapeadas as diversas ligações carbono-halogénio e carbono-oxigénio em ambiente alifático, olefínico e aromático sendo avaliada a capacidade de distinguir os diferentes tipos de ligação em causa. Relativamente à capacidade de distinção entre ligações carbono-azoto de aminas, iminas e nitrilos foi analisada a seguinte molécula:

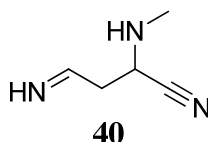


Figura III.10: Estrutura onde estão evidenciados diferentes tipos de ligação carbono-azoto.

No seguimento do estudo sobre a capacidade de distinção entre ligações carbono-azoto foi estudada uma molécula adicional:

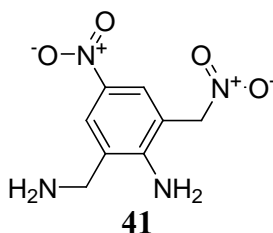


Figura III.11: Molécula onde estão evidenciadas ligações carbono-azoto em aminas e grupos nitro, tanto em ambiente alifático como aromático.

Neste caso foi estudada a capacidade dos diferentes mapas em estudo de separarem ligações carbono-azoto de aminas e grupos nitro em ambiente aromático e alifático. Para finalizar o estudo de capacidade de discriminação entre ligações carbono-azoto foi considerada ainda outra molécula:

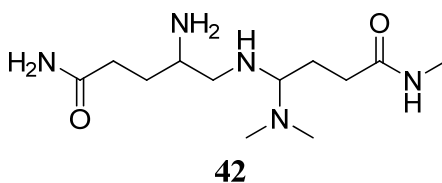


Figura III.12: Molécula apresentando diversas ligações carbono-azoto em aminas primárias, secundárias, terciárias e amidas.

Avaliou-se a capacidade de discriminação entre ligações carbono-azoto de amidas e aminas e entre aminas primárias, secundárias e terciárias. Relativamente à avaliação da capacidade de discriminação entre ligações carbono-oxigénio foi avaliada a seguinte estrutura:

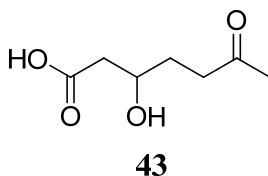


Figura III.13: Molécula apresentando diversos tipos de ligações carbono-oxigénio.

Foram mapeadas as diferentes ligações carbono-oxigénio (cetonas, álcoois e ácidos carboxílicos) em cada um dos mapas em comparação. Finalmente neste estudo de escolha do conjunto de descritores de ligações mais adequado para discriminação entre ligações semelhantes mas em contextos de grupos funcionais e ambientes químicos diferenciais foi considerado ainda o composto aromático seguinte:

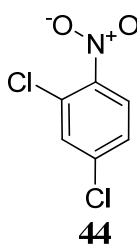


Figura III.14: Composto aromático substituído.

Relativamente à estrutura **44** foi estudada a capacidade em distinguir a diferentes ligações carbono-carbono aromáticas associadas a diferentes grupos substituinte incluídos os casos em que não estão associados grupos substituintes.

Relativamente a cada um dos estudos **a)**, **b)**, **c)** e **d)** foram seleccionados os seguintes conjuntos óptimos de descritores de ligações químicas:

a) Estudo do Mecanismo Reaccional de Compostos na Presença de butilamina

Foram seleccionados 3 conjuntos de descritores distintos: **i)** 27 descritores físico-químicos determinados pelo programa *cxcalc* pertencente ao ambiente Marvin, **ii)** 45 descritores topológicos extraídos do ficheiro de *output* gerado pelo programa *cxcalc*, **iii)** 59 descritores topológicos e físico-químicos extraídos do ficheiro de *output* gerado pelo programa *cxcalc*.

Tabela III.1-1: Descritores de ligações usados no estudo a) de mecanismo reaccional associado a determinado composto.

Descritor	Descrição	Tipo de descritor
1-HA	Se na ligação AB o átomo A é o hidrogénio	T
2-CA	Se na ligação AB o átomo A é o carbono	T
3-NA	Se na ligação AB o átomo A é o azoto	T
4-OA	Se na ligação AB o átomo A é o oxigénio	T
5-PA	Se na ligação AB o átomo A é o fósforo	T
6-SA	Se na ligação AB o átomo A é o enxofre	T
7-XA	Se na ligação AB o átomo A é halogénio	T
8-HB	Se na ligação AB o átomo B é o hidrogénio	T
9-CB	Se na ligação AB o átomo B é o carbono	T
10-NB	Se na ligação AB o átomo B é o azoto	T
11-OB	Se na ligação AB o átomo B é o oxigénio	T
12-PB	Se na ligação AB o átomo B é o fósforo	T
13-SB	Se na ligação AB o átomo B é o enxofre	T
14-XB	Se na ligação AB o átomo B é halogénio	T
15-HnA	O número de vizinhos do átomo A, de uma ligação AB, que são átomos de hidrogénio	T
16-CnA	O número de vizinhos do átomo A, de uma ligação AB, que são átomos de carbono	T
17-NnA	O número de vizinhos do átomo A, de uma ligação AB, que são átomos de azoto	T

T – Topológico.

Tabela III.1-2: Continuação da apresentação dos descritores de ligações usados no estudo a) de mecanismo reaccional associado a determinado composto.

Descritor	Descrição	Tipo de descritor
19-PnA	O número de vizinhos do átomo A, de uma ligação AB, que são átomos de fósforo	T
20-SnA	O número de vizinhos do átomo A, de uma ligação AB, que são átomos de enxofre	T
21-XnA	O número de vizinhos do átomo A, de uma ligação AB, que são halogéneos	T
22-Csp1_neigA	Numa ligação AB quantos carbonos sp são vizinhos ao átomo A;	T
23-Csp2_neigA	Numa ligação AB quantos carbonos sp ₂ são vizinhos ao átomo A	T
24-Csp3_neigA	Numa ligação AB quantos carbonos sp ₃ são vizinhos ao átomo A	T
25-HnB	O número de vizinhos do átomo B, de uma ligação AB, que são átomos de hidrogénio	T
26-CnB	O número de vizinhos do átomo B, de uma ligação AB, que são átomos de carbono	T
27-NnB	O número de vizinhos do átomo B, de uma ligação AB, que são átomos de azoto	T
28-OnB	O número de vizinhos do átomo B, de uma ligação AB, que são átomos de oxigénio	T
29-PnB	O número de vizinhos do átomo B, de uma ligação AB, que são átomos de fósforo	T
30-SnB	O número de vizinhos do átomo B, de uma ligação AB, que são átomos de enxofre	T
31-XnB	O número de vizinhos do átomo B, de uma ligação AB, que são halogéneos	T
32-Csp1_neigB	Numa ligação AB quantos carbonos sp são vizinhos ao átomo B;	T
33-Csp2_neigB	Numa ligação AB quantos carbonos sp ₂ são vizinhos ao átomo B	T
34-Csp3_neigB	Numa ligação AB quantos carbonos sp ₃ são vizinhos ao átomo B	T

T – Topológico.

Tabela III.1-3: Continuação da apresentação dos descritores de ligações usados no estudo a) de mecanismo reaccional associado a determinado composto.

Descritor	Descrição	Tipo de descritor
35-Boord	Ordem da ligação AB	T
36-aromaticaA	Se o átomo A da ligação AB faz parte de algum sistema aromático	T
37-aromaticB	Se o átomo B da ligação AB faz parte de algum sistema aromático	T
38-Csp1AA	Se o átomo A é um carbono sp	T
39-Csp2AA	Se o átomo A é um carbono sp ₂	T
40-Csp3AA	Se o átomo A é um carbono sp ₃	T
41-Csp1BB	Se o átomo B é um carbono sp	T
42-Csp2BB	Se o átomo B é um carbono sp ₂	T
43-Csp3BB	Se o átomo B é um carbono sp ₃	T
44-aro_neigA	O número de átomos vizinhos ao átomo A, numa ligação AB, que pertençam a algum sistema aromático	T
45-aro_neigB	O número de átomos vizinhos ao átomo B, numa ligação AB, que pertençam a algum sistema aromático	T
46-qpiA	Carga π do átomo A	FQ
47-qpiB	Carga π do átomo B	FQ
48-dqpi	Diferença entre a carga π do átomo A e B	FQ
49-qsigmaA	Carga σ do átomo A	FQ
50-qsigmaB	Carga σ do átomo B	FQ
51-dqsigma	Diferença de cargas entre o átomo A e o átomo B	FQ
52-qtotA	Carga total do átomo A	FQ
53-qtotB	Carga total do átomo B	FQ
54-dqtot	Diferença entre a carga total do átomo A e B	FQ
55-maxqtotA	Carga total máxima entre os vizinhos do átomo A	FQ
56-maxqtotB	Carga total máxima entre os vizinhos do átomo B	FQ
57-minqtotA	Carga total mínima entre os vizinhos do átomo A	FQ
58-minqtotB	Carga total mínima entre os vizinhos do átomo B	FQ
59-oenpiA	Electronegatividade de orbital π referente ao átomo A	FQ
60-oenpiB	Electronegatividade de orbital π referente ao átomo B	FQ
61-doenpi	Diferença de electronegatividades π entre átomos A e B	FQ
62-oensigmaA	Electronegatividade de orbital σ referente ao átomo A	FQ
63-oensigmaB	Electronegatividade de orbital σ referente a átomo B	FQ
64-donesigma	Diferença de electronegatividade σ entre os átomos A e B	FQ
65-maxpolA	Máximo de polarizabilidade entre os vizinhos do átomo A	FQ

T – Topológico, FQ – Físico-Químico.

Tabela III.1-4: Continuação da apresentação dos descritores de ligações usados no estudo a) de mecanismo reaccional associado a determinado composto.

Descritor	Descrição	Tipo de descritor
66-maxpolB	Máximo de polarizabilidades entre os vizinhos do átomo B	FQ
67-minpolA	Mínimo de polarizabilidades entre os vizinhos do átomo A	FQ
68-minpolB	Mínimo de polarizabilidades entre os átomos vizinhos do átomo B	FQ
69-pichgdensA	Densidade de carga π do átomo A	FQ
70-pichgdensB	Densidade de carga π de átomo B	FQ
71-totchgdensA	Densidade de carga total relativa ao átomo A	FQ
72-totchgdensB	Densidade de carga total relativa ao átomo B	FQ

FQ – Físico-Químico.

Na **Tabela III.1** estão representados os 45 descritores topológicos (1-45) que fazem parte do conjunto **ii)**, tal como os 27 descritores físico-químicos (46-72) que compõem o conjunto **i)**. Finalmente a **Tabela III.1** contem a selecção de 59 descritores topológicos e físico-químicos (1-47, 49, 50, 52, 53, 59, 60, 62, 63, 69-72) que compõem o conjunto **iii)** de descritores de ligação.

b) Estudo do Mecanismo Reaccional Associado a Reacções em que Intervenha a butilamina

Foi usado um conjunto de 59 descritores topológicos e físico-químicos para representar as ligações químicas, o mesmo conjunto de descritores assinalado na secção anterior (1-47, 49, 50, 52, 53, 59, 60, 62, 63, 69-72 - **Tabela III.1**).

c) Estudo de Reactividade de Compostos na Presença de butilamina

Foram seleccionadas 11 propriedades físico químicas de ligações calculadas pelo programa Petra (versão 4):

Tabela III.2: 11 descritores físico-químicos calculados pelo programa Petra e suas especificações.

Descritor	Descrição	Critério para escolha da representação da ligação
qpiA	Carga π pertencente ao átomo A da ligação A-B	Representação de ligação iniciada por átomo de carga total mais elevada
qpiB	Carga π pertencente ao átomo B da ligação A-B
qsigmaA	Carga σ pertencente ao átomo A da ligação A-B
qsigmaB	Carga σ pertencente ao átomo B da ligação A-B
qtotA	Carga total pertencente ao átomo A da ligação A-B
qtotB	Carga total pertencente ao átomo B da ligação A-B
oensigmaA	electronegatividade de orbital associada a ligação σ referente ao átomo A de ligação A-B
oensigmaB	electronegatividade de orbital associada a ligação σ referente ao átomo B de ligação A-B
SQIT	soma de todas cargas σ trocadas em todas iterações entre os átomos A e B da mesma ligação
B_STABRE	Estabilização por ressonância de carga positiva no átomo A e carga negativa no átomo B na quebra heterolítica de ligação A-B
B_STABRE-L	Estabilização por ressonância de carga positiva no átomo B e carga negativa no átomo A na quebra heterolítica de ligação A-B	Representação de ligação iniciada por átomo de carga total mais baixa

Cada ligação usada na obtenção de mapas de ligações foi representada pelas 11 propriedades físico-químicas de ligações expressas na **Tabela III.2**. Dez dessas onze propriedades basearam-se na representação de ligação iniciada pelo átomo de maior carga total. A propriedade B_STABRE-L foi calculada para a representação de ligação iniciada pelo átomo de carga total mais baixa.

d) Estudo de Reactividade de Compostos na Presença de Cianoborohidreto de Sódio

Foi seleccionado o mesmo conjunto de 11 propriedades físico-químicas usadas no estudo c).

Posteriormente, usando conjunto de ligações (seleccionado especificamente para cada um dos estudos a), b), c) e d)) representadas por conjuntos de descritores normalizados (específicos para cada um dos estudos a), b), c) e d)) foram treinadas redes de *Kohonen* de diferentes dimensões e características de forma a obter mapas de ligações. No caso dos estudos a) e b) foram construídos três conjuntos de cinco mapas, cada conjunto referente às 27 propriedades físico-químicas, 45 propriedades topológicas e 59 propriedades físico-químicas/topológicas de ligação representadas na **Tabela III.1**. Cada conjunto refere-se a 5 mapas de dimensões distintas: 10×10, 15×15, 20×20, 25×25 e 29×29.

Relativamente aos estudos c) e d) foram construídos seis mapas de dimensões diferenciais: 7×7, 10×10, 15×15, 20×20, 25×25 e 29×29.

III-2.2.2: Obtenção de Descritores MOLMAPs

As redes de *Kohonen* treinadas foram usadas para mapear as ligações químicas de cada composto presente numa determinada base de dados e assim obter para cada composto um perfil das ligações aí existentes (**Figura III.5**). O perfil das ligações activadas no mapa de *Kohonen* pode ser substituída por uma representação numérica (**Figura III.6**), reflectindo a abundância de determinado tipo de ligação nessa molécula, sendo assim obtida a representação MOLMAP de uma determinada molécula. Cada posição do MOLMAP corresponde a uma posição do mapa treinado (um determinado neurónio na rede de *Kohonen*). Quando uma ligação activa um determinado neurónio da rede treinada vai ser adicionado um determinado número a essa posição, no caso mais simples o valor 1 (tipo de MOLMAP 1_0 testado neste capítulo), em casos mais complexos, às posições correspondentes a neurónios vizinhos do neurónio activado também vai ser adicionado um valor (MOLMAPs do tipo 10_1 e 10_3 testados neste capítulo). Em MOLMAPs do tipo 10_1 e 10_3 à posição correspondente ao neurónio activado é adicionado o valor 10, aos neurónios vizinhos é adicionado o valor 1 ou 3 conforme o tipo de MOLMAP 10_1 ou 10_3 respectivamente. Cada MOLMAP obtido vai ser transformado num vector, concatenando colunas de cada mapa de *Kohonen* contendo a representação numérica do perfil das ligações de cada composto/reacção.

As especificidades associadas a cada estudo foram as seguintes:

a) Estudo do Mecanismo Reaccional de Compostos na Presença de butilamina

Em cada rede de *Kohonen* treinada para este estudo foram mapeadas todas as ligações dos 1942 compostos (971 reagentes mais os respectivos 971 produtos) presentes na base de dados de reacções em que intervém a butilamina. Resultaram os seguintes tipos de MOLMAPs, divididos em três blocos de acordo com o conjunto de descritores usado para representar cada ligação química no treino e mapeamento:

Tabela III.3: Tipos e dimensões de MOLMAPs usados no estudo de mecanismo reaccional de composto. Em todos os casos foram usados descritores do tipo 10_3.

Dimensões do MOLMAP	Número de descritores correspondentes	Descritores de ligações químicas
10×10	100	27 descritores físico-químicos (Marvin 4.1.4)
15×15	225	*****
20×20	400	*****
25×25	625	*****
29×29	841	*****
10×10	100	45 descritores topológicos (Marvin 4.1.4)
15×15	225	*****
20×20	400	*****
25×25	625	*****
29×29	841	*****
10×10	100	59 descritores físico-químicos / topológicos (Marvin 4.1.4)
15×15	225	*****
20×20	400	*****
25×25	625	*****
29×29	841	*****

b) Estudo do Mecanismo Reaccional associado a Reacções em que Intervenha a butilamina

Neste estudo foram mapeados, em redes de *Kohonen*, 971 reagentes, respectivos 971 produtos e butilamina presentes na base de dados de reacções em que intervém a butilamina, sendo obtidos MOLMAPs para cada um dos compostos intervenientes. Posteriormente foi calculado MOLMAP de

cada reacção subtraindo ao MOLMAP de um determinado produto, os MOLMAPs do reagente respectivo e butilamina. Os MOLMAPs de reacção estudados possuem as seguintes características:

Tabela III.4: Tipos e dimensões de MOLMAPs usados no estudo de mecanismo reaccional associado a uma determinada reacção.

Dimensões do MOLMAP	Número de descritores correspondentes
10×10	100
15×15	225
20×20	400
25×25	625
29×29	841

Para cada dimensão foram considerados dois tipos de MOLMAP 1_0 e 10_3. Todos os MOLMAPs utilizados basiam-se em 59 descritores físico-químicos/topológicos.

c) Estudo de Reactividade de Compostos na Presença de butilamina

Em cada rede de *Kohonen* treinada para este estudo foram mapeadas todas as ligações dos 1942 compostos (971 reagentes mais os respectivos 971 produtos) presentes na base de dados de reacções em que intervém a butilamina. Resultaram os seguintes tipos de MOLMAPs:

Tabela III.5: Dimensões de MOLMAP usados no estudo de reactividade na presença de butilamina.

Dimensões do MOLMAP	Número de descritores correspondentes
7×7	49
10×10	100
15×15	225
20×20	400
25×25	625
29×29	841

Para cada dimensão foram considerados dois tipos de MOLMAP 1_0 e 10_1. Todos os MOLMAPs utilizados basiam-se em 11 descritores físico-químicos.

d) Estudo de Reactividade de Compostos na Presença de Cianoborohidreto de Sódio

Todas as ligações de cada um dos 728 compostos (364 reagentes e respectivos 364 produtos) provenientes da base de dados de reacções onde está envolvido o cianoborohidreto de sódio foram submetidas às redes de *Kohonen* treinadas, usadas neste estudo, resultando nos seguintes tipos de MOLMAP:

Tabela III.6: Dimensões de MOLMAPs usados no estudo de reactividade na presença de cianoborohidreto de sódio.

Dimensões do MOLMAP	Número de descritores correspondentes
7×7	49
10×10	100
15×15	225
20×20	400
25×25	625
29×29	841

Para cada dimensão foram considerados dois tipos de MOLMAP 1_0 e 10_1. Todos os MOLMAPs utilizados basiam-se em 11 descritores físico-químicos.

III-2.3: Conjuntos de Dados Derivados de Bases de Dados Reaccionais

Depois de obter MOLMAPs para compostos e reacções de bases de dados seleccionadas transformou-se a informação obtida em informação útil para construir, validar e testar os modelos de reactividade e mecanismo reaccional estudados neste capítulo. Segundo o tipo de modelo considerado **a)**, **b)**, **c)** ou **d)** foram seguidas estratégias diferentes.

No caso dos modelos de previsão de reactividade **c)** e **d)** é necessário informação de compostos que reagem, mas também de compostos que não reagem, e esta última informação não estava presente em nenhuma das bases de dados consideradas. Para extrair essa informação das bases de dados foi desenvolvido um procedimento específico.

a) Estudo do Mecanismo Reaccional de Compostos na Presença de butilamina

Partindo de conjunto de 971 reacções em que intervém a butilamina, foram calculados os respectivos MOLMAPs associados a cada reacção subtraindo ao MOLMAP de um determinado produto, os MOLMAPs do reagente respectivo e butilamina. Os MOLMAPs usados eram de dimensões 7X7 (49 descritores) do tipo 10_3 usando 59 descritores topológicos/físico-químicos

extraídos do ficheiro de *output* gerado pelo ambiente Marvin. As 971 reacções assim representadas foram usadas para treinar uma rede neuronal de *Kohonen* de dimensão 25X25, 49 camadas de *input* e treinada com 100 *epochs*. Nessa rede de *Kohonen* treinada foram mapeadas as 971 reacções consideradas. A partir desse mapeamento foram seleccionadas 194 reacções de forma a cobrir o espaço reaccional pretendido. Foram assim considerados 6 tipos de reacções diferenciados, de acordo com estudo de interpretação mecanística de Aptula *et al.*¹ a) formação de bases de Schiff, b) reacção de grupo acil, c) Substituição nucleofílica, d) Substituição nucleofílica aromática, e) Adição de Michael, f) Reacções cujos reagentes não apresentam características electrofílicas evidentes. As 194 reacções foram distribuídas por 2 conjuntos, um conjunto de treino de 145 reacções e um conjunto de validação de 49 reacções; adicionalmente foram geradas 45 reacções que foram incorporadas no conjunto de teste. Relativamente a estes 3 conjuntos, foram separados os reagentes dos respectivos produtos, ficando os 3 conjuntos representados exclusivamente por MOLMAPs de reagentes e respectiva classe reaccional. As características de cada conjunto de dados aparecem expressas na seguinte tabela (**Tabela III.7**):

Tabela III.7: Composição dos conjuntos de treino, validação e teste relativamente aos diferentes tipos de mecanismos reaccionais.

		Conjunto de treino	Conjunto de validação	Conjunto de teste
Número objectos total		145	49	45
Número de objectos por classe reaccional	Reagentes usados na formação de bases de Schiff a)	26	9	12
	Reagentes contendo grupos acilo reactivos b)	19	7	8
	Reagentes que reagem segundo mecanismo S ₂ N c)	54	18	9
	Reagentes que reagem segundo mecanismo S _N Ar d)	20	6	11
	Reagentes que reagem segundo adição de Michael e)	8	2	5
	Reagente com características electrófilas não evidentes f)	18	7	0

b) Estudo do Mecanismo Reaccional Associado a Reacções em que Intervenha a butilamina

Os conjuntos de treino, validação e teste são idênticos aos do estudo **a)** mas neste caso os objectos em estudo são reacções e não reagentes como no estudo **a)**. Neste caso cada objecto (reacção) está representado pelo respectivo MOLMAP e classe reaccional associada.

c) Estudo de Reactividade de Compostos na Presença de butilamina

Partindo das 971 reacções em que intervém a butilamina (**Esquema III.7**),



Esquema III.7: Representação genérica de uma reacção em que intervém a butilamina.

seleccionadas de acordo com critérios da secção **III-2.1.2**, estas foram fraccionadas em duas partes, uma de 922 reacções, usada como base para o conjunto de treino, e outra de 49 reacções usada como base para conjuntos de validação e de previsão.

Conjunto de Treino

Relativamente ao conjunto de 922 reacções usadas como base para o conjunto de treino e tendo em conta que cada reacção estava representada por MOLMAPs de reagente, butilamina e respectivo produto, de acordo com protocolo descrito na secção **III-2.2.2**, seguiram-se dois procedimentos distintos tendo em conta a derivação de informação negativa (compostos que não reagem):

c1) R-P

Comparou-se directamente MOLMAP do reagente (R) ao MOLMAP do respectivo produto (P). A comparação foi feita entre posições equivalentes de MOLMAPs do mesmo tipo (**Figura III.15**).

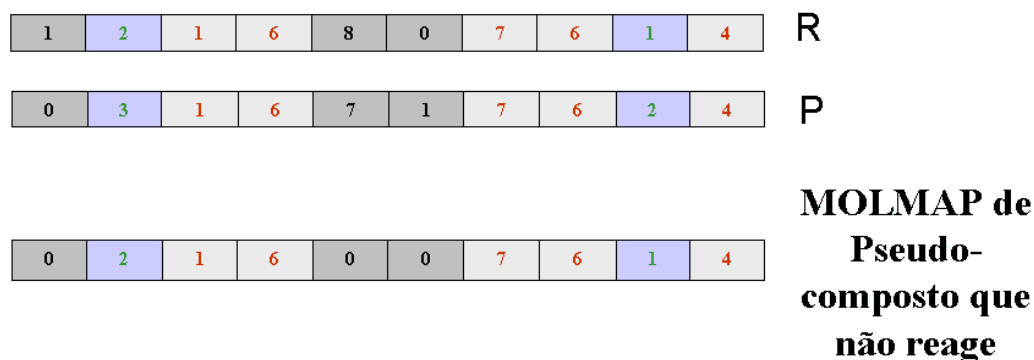


Figura III.15: Esquema de derivação de MOLMAP de pseudo-composto que não reage.

A comparação consistia em subtrair ao MOLMAPs de (R) o MOLMAP de (P). Se para determinada posição a diferença entre (R) e (P) fosse maior que “0” significava que determinado tipo de ligação desaparecia na conversão de reagente para produto logo a essa posição é atribuído o valor “0” no MOLMAP resultante (**Figura III.15**). No caso da diferença entre (R) e (P) ser igual a zero para determinada posição significa que o tipo de ligação correspondente não sofre alteração durante a reacção, nesse caso vai ser atribuído o valor do reagente (R) ao MOLMAP resultante (**Figura III.15**). No caso da diferença entre (R) e (P), para determinada posição, ser menor que “0” significa que se formou uma ligação nova no produto que não existia no reagente, nesse caso vai ser atribuído o valor de (R) no MOLMAP resultante (**Figura III.15**). Desta forma o conjunto de treino vai estar constituído por 1844 objectos, 922 MOLMAPs reagentes que “reagem” provenientes directamente da base de dados reaccional e 922 MOLMAPs de pseudocompostos da classe “não reage”.

c2) R-P*

Este procedimento é essencialmente idêntico ao anterior **c1)** a única alteração reside na definição do MOLMAP do produto usado na comparação com o MOLMAP do reagente. Antes dessa comparação, ao MOLMAP do produto (P) é subtraído o MOLMAP da butilamina sendo obtido MOLMAP de produto (P*), descontando as ligações da butilamina. O MOLMAP do produto (P*) foi comparado ao MOLMAP do respectivo reagente (R) da mesma forma descrita em **c1)**. Esta alteração deveu-se à preocupação em evitar que eventuais alterações em ligações de amins do reagente fossem eclipsadas pela presença do fragmento butilamina no produto. Assim, à semelhança do procedimento **c1)**, o conjunto de treino vai estar composto por 1844 objectos, metade “reage” a outra metade “não reage”.

Conjuntos de Previsão e Validação

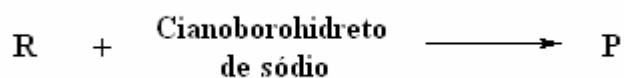
Relativamente ao conjunto de 49 reacções usadas como base para o conjunto de validação e de teste, inspeccionou-se cada reagente e detectaram-se manualmente as ligações com potencial para poder reagir com a butilamina. Essas ligações foram apagadas e/ou alteradas obtendo-se assim 49 compostos que se consideram não reagir com butilamina. Obteve-se assim um conjunto de 98 compostos (reagentes), constituído por 49 compostos da classe “reage” - *c* (reagentes provenientes do conjunto de 49 reacções usadas como base), e por 49 compostos da classe “não reage” - *a* (derivados dos primeiros apagando/modificando ligações com potencial para reagir). Este conjunto de 98 compostos foi distribuído de forma aleatória e equilibrada (relativamente à composição das duas classes, “reage” e “não reage”) por dois conjuntos:

- 1) Conjunto de validação de 49 compostos (25 compostos da classe “reage” - *c* e 24 compostos da classe “não reage” - *a*) representado cada um pelo seu respectivo MOLMAP e classe.
- 2) Conjunto de previsão de 49 compostos (24 compostos da classe “reage” - *c* e 25 compostos da classe “não reage” - *a*) representa cada um pelo seu respectivo MOLMAP e classe.

De forma a testar a capacidade do modelo de reactividade em prever o potencial de um composto em provocar dermatite de contacto alérgica foi usado um conjunto de 32 compostos testados *in vitro* num estudo prévio.³

d) Estudo de Reactividade de Compostos na Presença de Cianoborohidreto de Sódio

Partindo de conjunto de 364 reacções em que intervém o cianoborohidreto de sódio (**Esquema III.8**):



Esquema III.8: Representação genérica de uma reacção em que intervém o cianoborohidreto de sódio.

seleccionadas de acordo com critérios da secção **III-2.1.2**, estas foram fraccionadas em duas partes, uma de 319 reacções, usada como base para o conjunto de treino, e outra de 45 reacções usada como base para conjuntos de validação e de previsão.

Conjunto de Treino

Relativamente ao conjunto de 319 reacções usadas como base para o conjunto de treino e tendo em conta que cada reacção estava representada por MOLMAPs de reagente e respectivo produto (de acordo com procedimento descrito na secção **III-2.2.2.**), foi derivada informação negativa relativa a pseudocompostos que não reagem de acordo com procedimento **c1**) descrito nesta mesma secção **III-2.3.**. Assim, o conjunto de treino ficou constituído por 638 objectos, 319 MOLMAPs de compostos da classe “reage” - c provenientes directamente da base de dados reaccional em que intervém o cianoborohidreto de sódio e 319 MOLMAPs de pseudocompostos da classe “não reage” - a derivados dos primeiros 319 MOLMAPs relativos a compostos presentes na base de dados reaccional (319 reacções).

Conjuntos de Previsão e Validação

Relativamente ao conjunto de 45 reacções como base para constituição de conjuntos de validação e de previsão, olhou-se para cada reagente e observou-se quais as ligações com potencial para poder reagir com o cianoborohidreto de sódio, essas ligações foram apagadas e/ou alteradas obtendo-se assim 45 compostos que não iriam reagir na presença de cianoborohidreto de sódio. Desta forma obteve-se um conjunto de 90 compostos (reagentes), constituído por 45 compostos da classe “reage” - c (reagentes provenientes do conjunto de 45 reacções base), e por 45 compostos da classe “não reage” - a (derivados dos primeiros apagando e/ou modificando ligações com potencial para reagir). Este conjunto de 90 compostos foi distribuído de forma aleatória e equilibrada (relativamente à distribuição de classes) por dois conjuntos de objectos:

- 1) Conjunto de validação constituído por 44 compostos provenientes do conjunto de 90 compostos descrito acima, adicionado de 2 compostos introduzidos especialmente neste conjunto, somando no total 46 compostos (24 compostos da classe “reage” - c e 22 compostos da classe “não reage” - a).
- 2) Conjunto de previsão constituído por 46 compostos provenientes de conjunto de 90 compostos descritos acima, adicionado de 2 compostos introduzidos especialmente neste conjunto, somando no total 48 compostos (25 compostos da classe “reage” - c e 23 compostos da classe “não reage” - c). Cada um dos compostos está representado por um MOLMAP e classe respectiva.

Ainda a referir que os 4 compostos (2 + 2) distribuídos por conjunto de validação e de previsão continham ligações carbono-halogéneo e funcionalidade oxi-metilamina (**Figura 16**), grupos que vão

reagir na presença de cianoborohidreto de sódio mas que não estavam muito representados na base de dados original. Cada um dos compostos foi representado por um MOLMAP e classe respectiva.

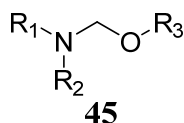


Figura III.16: Funcionalidade oxi-metilamina.

Foi usado um conjunto adicional de 25 compostos de forma a testar a capacidade do modelo de reactividade em discriminar funcionalidades específicas. Iminas, aldeídos, cetonas, ácidos carboxílicos, ésteres e imidas são as funcionalidades presentes neste conjunto de 25 compostos.

III-2.4: Modelos Construídos

Com os conjuntos de dados seleccionados/derivados foram construídos, validados e testados uma série de modelos referentes a cada um dos estudos **a)**, **b)**, **c)** e **d)** abordados neste capítulo.

III-2.4.1. a): Classificação de Compostos de Acordo com o Mecanismo de Reacção com a butilamina

Neste estudo procurou-se construir um modelo para classificar compostos que reagem com butilamina, em termos do mecanismo reaccional. Foram estudadas uma série de variáveis, começando pelos conjuntos de propriedades associadas a cada ligação química: 27 variáveis físico-químicas, 45 características topológicas e 59 características topológicas e físico-químicas. Posteriormente estudou-se a dimensão e tipo de MOLMAP mais adequado. Nestes estudos foram usadas *Random Forest* como método de aprendizagem automático. Relativamente ao uso de algoritmos sensíveis ao número de descritores e sua inter-correlação, como é o caso de redes neuronais de *Kohonen*, foram usados diversos métodos de selecção/contabilização de descritores MOLMAP. Os modelos obtidos foram comparados através das previsões obtidas para um conjunto de validação de 49 compostos. Os modelos mais adequados foram testados usando um conjunto de teste de 45 compostos. Nestes estudos as variáveis de *input* são os MOLMAPs dos compostos e o mecanismo é o objectivo a obter (*output*).

Uso de 27 Variáveis Físico-Químicas

Foi usado o algoritmo *Random Forest* (1000 árvores, $mtry = 25$) na obtenção de modelos de classificação segundo mecanismo reaccional. Neste estudo variou-se a dimensão de MOLMAPs treinados com 27 propriedades físico-químicas, tal como aparece descrito na **Tabela III.3**.

Uso de 45 Variáveis Topológicas

À semelhança do estudo anterior foi usado o algoritmo *Random Forest* (1000 árvores, $mtry = 25$) na obtenção de modelos de classificação, neste caso também se variou a dimensão de MOLMAPs obtidos através de mapeamento segundo 45 características topológicas, tal como aparece descrito na **Tabela III.3**.

Uso de 59 Variáveis Topológicas e Físico-Químicas

Este estudo implicou o uso dos algoritmos *Random Forest* e redes neuronais de *Kohonen*. Algoritmos com características muito diferentes relativamente à informação proporcionada e à sensibilidade ao número e correlação dos descritores MOLMAP utilizados. Tendo em conta esse diferencial foram usados diferentes métodos na construção destes dois tipos de modelos:

Modelos *Random Forest*

Neste estudo usou-se como parâmetros 1000 árvores variando-se o tipo e dimensão dos MOLMAPs tal como aparece descrito na **Tabela III.3**, variou-se também o valor de $mtry$ (10, 25, 50, 75, 100), sendo seleccionado $mtry = 10$, pois permitiu concordância para conjunto de validação de 49 compostos de mais elevada.

Modelo Árvore de Decisão

Foi construído um modelo Árvore de Decisão usando parâmetros de referência. O modelo final está composto por 8 regras lógicas (8 descritores).

Modelos Redes Neurais de *Kohonen*

Para este estudo foi necessária uma selecção prévia dos componentes MOLMAP, sendo usados como ponto de partida MOLMAPs 15×15 (225 descritores) do tipo 10_3. Os descritores usados foram normalizados através de normalização Z (**Equação III.3**). Foram usados 3 métodos de selecção distintos:

1. Neste método foram usados algoritmos genéticos para seleccionar descritores MOLMAPs, cujos parâmetros consistiam em 200 gerações cada geração contendo uma população 200 elementos, durante o processo de selecção foram usadas redes neurais de *Counter-Propagation* 13×13 com 6 camadas de *output*, correspondendo cada camada a um tipo de mecanismo reaccional, a ocorrência de determinado mecanismo foi assinalada com o valor “1” e a sua ausência era representada pelo valor “0”. O critério de selecção de descritores baseava-se no valor de RMS, relativo às previsões para conjunto de treino, obtido por rede neuronal de *Counter-Propagation* treinada com determinado conjunto gerado de descritores. Foi assim seleccionado um conjunto de 32 descritores dos 225 componentes MOLMAP iniciais. Esse conjunto de descritores foi usado no treino de uma rede neuronal de *Kohonen* 13×13 treinada em 100 *epochs*.
2. Relativamente a este método foram seleccionadas componentes de MOLMAPs de reacções cujos valores mínimos (tendo em conta as 145 reacções consideradas na construção do conjunto de treino) eram negativos, pois significava o desaparecimento de determinada ligação na conversão de reagentes em produtos. Desta forma foram seleccionados 68 descritores MOLMAP de reagentes correspondendo às posições seleccionadas nos MOLMAPs de reacção; estes descritores foram usados no treino de uma rede neuronal de *Kohonen* 13×13 treinada em 100 *epochs*.
3. Este método baseou-se no treino de uma rede neuronal de *Kohonen* de dimensões 13×13 treinada em 100 *epochs*. Na fase de treino cada objecto estava representado por MOLMAPs de reagente (15×15 – 225 descritores) e MOLMAPs de reacção (15×15 – 225 descritores). Posteriormente à fase de treino, foram apagadas as camadas da rede neuronal de *Kohonen* correspondente a MOLMAPs de reacções, permanecendo as 225 camadas correspondentes a MOLMAPs de reagentes.

Os diferentes métodos de selecção/contabilização de descritores foram comparados entre si através da capacidade de previsão para conjunto de treino de 145 compostos e conjunto de validação de 49 compostos.

III-2.4.2 b): Classificação de reacções com butilamina

Neste estudo usou-se como parâmetros 1000 árvores variando-se o tipo e dimensão dos MOLMAPs tal como aparece descrito na **Tabela III.4**, variou-se também o valor de *mtry* (10, 25, 50, 75, 100), sendo seleccionado *mtry* = 25, o valor base usado neste estudo, já que outros valores de *mtry* não provocaram incremento na concordância para conjunto de validação de 49 compostos. Os MOLMAPs de reacção utilizados baseiam-se no mapeamento de acordo com 59 variáveis físico-químicas e topológicas de ligações químicas, as mesmas que foram utilizadas no estudo **a)**.

III-2.4.3 c): Estudo de Reactividade de Compostos na Presença de butilamina

Foram construídos e comparados uma série de modelos, cada modelo com características específicas quanto à dimensão e tipo de MOLMAP, método para derivação de informação negativa, método de aprendizagem automático e parâmetros específicos associados ao algoritmo de aprendizagem automático. Todos estes parâmetros foram optimizados olhando para as previsões atribuídas ao conjunto de validação. O melhor modelo foi testado usando um conjunto de previsão e um conjunto de compostos testados *in vitro* na presença de péptido contendo lisina.

Os métodos para obtenção de informação negativa (MOLMAP de pseudocompostos que não reagem) estão explicados em **III-2.3**. Aparecem agrupados os modelos construídos tendo em conta o método de aprendizagem automático utilizado:

Random Forest. Usando como parâmetros base, número de árvores = 1000 e valor de *mtry* = 25 foram construídos uma série de modelos *Random Forest* usando o ambiente R versão 2.4.0⁴³ através da biblioteca randomForest. Nestes estudos as variáveis de *input* são os MOLMAPs dos compostos e a classe (reage ou não reage) é o objectivo a obter (*output*). Variou-se as dimensões e tipo de MOLMAP para representar cada composto tal como se variou o método para derivação de informação negativa relativamente a pseudocompostos que não reagem. Os modelos construídos e as suas características aparecem condensados na **Tabela III.8**:

Tabela III.8: Características dos modelos *Random Forest* construídos, tendo em conta o tipo de MOLMAP, suas dimensões e método de obtenção de pseudocompostos que não reagem.

Tipo de MOLMAP	Método de obtenção de pseudocompostos que não reagem: c1) R-P* - Comparando MOLMAP de Reagente (R) à diferença entre MOLMAPs de Produto e butilamina (P*)		Método de obtenção de pseudocompostos que não reagem: c2) R-P - Comparando MOLMAP de Reagente (R) ao MOLMAP de Produto (P)	
	10 1	1 0	10 1	1 0
Dimensões do MOLMAP	29×29	29×29	29×29	29×29
	25×25	25×25	25×25	25×25
	20×20	20×20	20×20	20×20
	15×15	15×15	15×15	15×15
	10×10	10×10	10×10	10×10
	7×7	7×7	7×7	7×7

Foi ainda optimizado o parâmetro *mtry* da *Random Forest* variando-o de 25 a 300. O modelo com as características seleccionadas (MOLMAP 20×20, tipo 1_0, método c1) de obtenção de pseudocompostos que não reagem *mtry* = 50 e 1000 árvores) foi usado na previsão de reactividade para um conjunto independente de previsão de 49 compostos. Adicionalmente este modelo foi avaliado relativamente à capacidade de previsão do potencial de determinado composto em provocar dermatite de contacto alérgica, recorrendo a um conjunto de 32 compostos cuja reactividade na presença de péptido contendo lisina tinha sido testada previamente.³

O modelo foi validado por randomização da variável dependente (y-randomization). Dispondo de forma aleatória a classe de reactividade associada a cada objecto do conjunto de treino, ainda que mantendo a mesma proporção de objectos que reagem e não reagem, foi treinado modelo que possuía exactamente as mesmas características do modelo seleccionado. O modelo obtido foi avaliado relativamente à capacidade de previsão dos conjuntos de validação e de teste.

Árvores de Decisão. Complementarmente aos modelos *Random Forest* foram construídos modelos baseados em Árvores de Decisão que auxiliaram na escolha do melhor tipo de MOLMAP, método de geração de pseudocompostos que não reagem e interpretação de descritores MOLMAP. Foi usado o ambiente R versão 2.4.0 com a biblioteca *rpart*, os parâmetros usados na construção de cada modelo árvore de decisão eram os valores pré-definidos.

Seleção de Descritores. Foi realizada selecção de descritores para métodos de aprendizagem automática sensíveis ao número, correlação e qualidade dos descritores (redes neuronais de *Kohonen* e *Support Vector Machines*). Partindo de mapa de ligações baseado em rede de

Kohonen de dimensões 29×29 foram aí mapeadas todas as ligações dos 922 reagentes e respectivos produtos seleccionados (secção III-2.3 c). Cada ligação possuía uma etiqueta relativamente à sua pertença a um reagente (valor -1) ou um produto (valor 1). Nesse mapeamento foram seleccionados aqueles neurónios (posições do mapa) que possuíam maior proporção de ligações de reagentes. Pretendeu-se assim identificar componentes do MOLMAP associados a ligações que desaparecem dos reagentes e que corresponderiam a centros reactivos.

Redes Neurais de Kohonen. Aqui foram usadas redes neuronais de *Kohonen* para classificar compostos. Estas redes são independentes das redes usadas para gerar os MOLMAPs. São usadas em vez de *Random Forests* ou Árvores de Decisão referidas acima nesta secção. Nestes estudos as variáveis de *input* são os MOLMAPs dos compostos. Depois do treino não-supervisionado os neurónios são classificados (“pintados”) de acordo com a classe (reage ou não reage) dos compostos que os activaram. É esta classificação dos neurónios que serve para classificar compostos novos.

Variou-se o número de descritores MOLMAP 29×29 seleccionados (64 e todos os 841 descritores), o tipo de MOLMAP considerado (1_0 e 10_1), tal como as dimensões das redes de *Kohonen* treinadas (10×10, 15×15, 20×20, 25×25 e 29×29). Todos os descritores foram normalizados usando normalização Z (**Equação III.3**). A qualidade dos modelos foi avaliada através da percentagem previsões acertadas para um conjunto de validação de 49 compostos.

Support Vector Machines. Também neste estudo as variáveis de *input* são os MOLMAPs dos compostos e a classe (reage ou não reage) é o objectivo a obter (*output*). Usando a mesma selecção de 64 descritores MOLMAP (29×29) variou-se o parâmetro *C* de complexidade associado aos modelos *Support Vector Machines* construídos (valores de *C* testados 1, 2, 4, 8, 16, 32 e 64), foram usados MOLMAPs do tipo 10_1 normalizados internamente pelo método das *Support Vector Machines*. A qualidade dos modelos foi avaliada através da percentagem de previsões acertadas para um conjunto de validação de 49 compostos.

III-2.4.4 d): Estudo de Reactividade de Compostos na Presença de Cianoborohidreto de Sódio

À semelhança do estudo c) de reactividade na presença de butilamina, neste estudo foram construídos e comparados uma série de modelos, cada modelo possui um conjunto de características específico desde a dimensão e tipo de MOLMAP, método de aprendizagem automático e parâmetros associados. A optimização de um modelo de classificação de reactividade baseou-se nas previsões obtidas para um conjunto de validação de 46 compostos. O modelo optimizado foi testado através de dois conjuntos de teste (48 e 25 compostos). Os métodos de aprendizagem utilizados na construção dos modelos são *Random Forest* e *Árvore de Decisão*, em cada caso serão focadas as características testadas na obtenção de um modelo optimizado.

Modelos *Random Forest*

Foram construídos uma série de modelos RF usando como parâmetros base 1000, como número de árvores construídas, e 25 como valor de *mtry*. Inicialmente variou-se as dimensões e tipo de MOLMAPs.

Tabela III.9: Características dos MOLMAPs usadas na construção de modelos *Random Forest*, variando o tipo e as dimensões dos MOLMAP.

Método de obtenção de pseudocompostos que não reagem: c2) R-P - Comparando MOLMAP de Reagente (R) ao MOLMAP de Produto (P)		
Tipo de MOLMAP	10_1	1_0
Dimensões do MOLMAP	29×29	29×29
	25×25	25×25
	20×20	20×20
	15×15	15×15
	10×10	10×10
	7×7	7×7

Depois de seleccionar as características de MOLMAP óptimas, foi feita uma optimização mais fina estudando a influência do parâmetro *mtry*, valores compreendidos na gama (15-225) foram testados. Foi encontrado o melhor modelo comparando as previsões obtidas para o conjunto de validação. O modelo seleccionado para além de ter sido testado na previsão de reactividade para dois conjuntos independentes de teste, foi validado num ensaio descrito em seguida.

Validação por Aleatoriedade de Classe de Reactividade

Neste ensaio foi disposta de forma aleatória a classe de reactividade associada a cada objecto do conjunto de treino ainda que mantendo a mesma proporção de objectos que reagem e não reagem, seguidamente foi treinado modelo que possuía exactamente as mesmas características do modelo seleccionado, o modelo obtido foi avaliado relativamente à capacidade de previsão relativamente a conjuntos de validação e de teste.

Modelos Árvore de Decisão

De forma complementar os modelos *Random Forest*, foram construídos modelos baseados no algoritmo da árvore de decisão. Usando parâmetros pré-definidos associados a este algoritmos foram construídos uma série de modelos em que se variou as dimensões e tipo de MOLMAP tal como foi descrito na **Tabela III.6**.

III-2.5: Medidas de Qualidade dos Modelos de Previsão

Definições de concordância, k , λ_B , predictibilidade e sensibilidade aparecem descritas na secção II-2.4 b) – Metodologia Computacional

Para problema de duas classes, negativa (compostos não reactivos) e positiva (compostos reactivos) a **especificidade** está definida matematicamente por:

$$Especificidade = \frac{Negativos_Verdadeiros}{Negativos_Verdadeiros + Falsos_Positivos}$$

Equação III.4

Por sua vez, a **sensibilidade** está definida matematicamente por:

$$Sensibilidade = \frac{Positivos_Verdadeiros}{Positivos_Verdadeiros + Falsos_Negativos}$$

Equação III.5

III-2.6: Interpretação de Descritores MOLMAP

De forma transversal aos quatro estudos descritos neste capítulo foi realizada interpretação dos descritores MOLMAP mais relevantes, seleccionados através de métodos de aprendizagem automática. Cada descritor MOLMAP corresponde a um ou vários tipos de ligação, correspondendo a um determinado perfil de propriedades físico-químicas de ligação. De maneira a associar um determinado descritor MOLMAP a tipos de ligação, o seguinte procedimento é seguido:

- 1) Em mapa auto-organizativo de determinadas dimensões, usado na obtenção de descritores MOLMAP, foi mapeado conjunto de ligações químicas usadas para o treinar. Nesse mapeamento cada ligação química vai ser colocada em determinado neurónio do mapa auto-organizativo.
- 2) Determinado descritor relevante vai corresponder a determinado neurónio em mapa auto-organizativo. Nesse mapa a atribuição do número do descritor vai ser feita em ordem

crescente, começando na primeira coluna do extremo esquerdo do mapa, percorrendo todas as linhas dessa coluna desde o extremo superior até ao extremo inferior. A atribuição termina na coluna do extremo direito.

- 3) Em determinado neurónio correspondendo a determinado descritor relevante vão estar mapeadas determinadas ligações. Cada ligação vai estar associada a uma determinada molécula da base de dados original, tal como à localização dos dois átomos concretos pertencentes a essa ligação.
- 4) Assim, cada descritor vai estar associado a determinado(s) tipo(s) de ligação referentes aos átomos envolvidos, ordem de ligação e ambiente químico.

III-3: Resultados e Discussão

III-3.1 a): Classificação de Compostos quanto ao Mecanismo da Reacção com butilamina

Nesta secção pretende-se construir modelos que classifiquem (prevejam) o mecanismo reaccional associado a determinado composto colocado a reagir na presença de butilamina, nucleófilo modelo. Ao construir estes modelos pretende-se testar se os descritores MOLMAP têm capacidade para evidenciar padrões reaccionais associados à estrutura de determinado composto, e no caso de evidenciarem essa capacidade descobrir quais as características estruturais determinantes para um composto seguir determinado mecanismo reaccional na presença de nucleófilo e não outro mecanismo qualquer. Este estudo tem ainda como objectivo prever como determinado composto químico, com potencial para provocar dermatite de contacto alérgica, poderá reagir com os aminoácidos nucleófilos presentes na pele.

Os seis tipos de mecanismos reaccionais considerados neste estudo são: **i)** formação de bases de Schiff (classe a), **ii)** reacção de acilação (classe b), **iii)** substituição nucleofílica (classe c), **iv)** substituição nucleofílica aromática (S_NAr) (classe d), **v)** adição de Michael (classe e) e **vi)** Outros tipos de reacção (classe f), de acordo com **Esquemas III.1-III.6** da Introdução Teórica.

A selecção dos melhores conjuntos de descritores de ligação química é um dos passos mais importantes e mais delicados na obtenção de modelos que tenham a capacidade de discriminar os diferentes mecanismos reaccionais envolvidos. A selecção dos melhores conjuntos de descritores de ligação química baseou-se na capacidade de discriminação de ligações químicas envolvidas em

ambientes químicos diversos. Concretamente, para um conjunto genérico de propriedades de ligação, depois de normalizadas, foram usadas para treinar um mapa auto-organizativo (SOM) 29x29 onde foram mapeados diversos conjuntos de ligações, cada conjunto continha ligações iguais ou semelhantes mas com ambiente químico diferente. Assim para cada conjunto específico de ligações avaliou-se a capacidade de discriminação segundo o ambiente químico envolvido em cada ligação. Por exemplo, um conjunto específico de ligações químicas envolve a ligação C=O de carbonilos em diferentes ambientes químicos como aldeídos, cetonas, amidas, ácidos carboxílicos entre outras funcionalidades. Segundo este critério de capacidade de discriminação foram seleccionados três conjuntos de descritores de ligação química: 59 descritores topológicos e físico-químicos, 45 descritores exclusivamente topológicos e 27 descritores físico-químicos. Nesta discussão será focado em detalhe a capacidade de discriminação proporcionado por cada conjunto de descritores (59, 45 e 27 descritores).

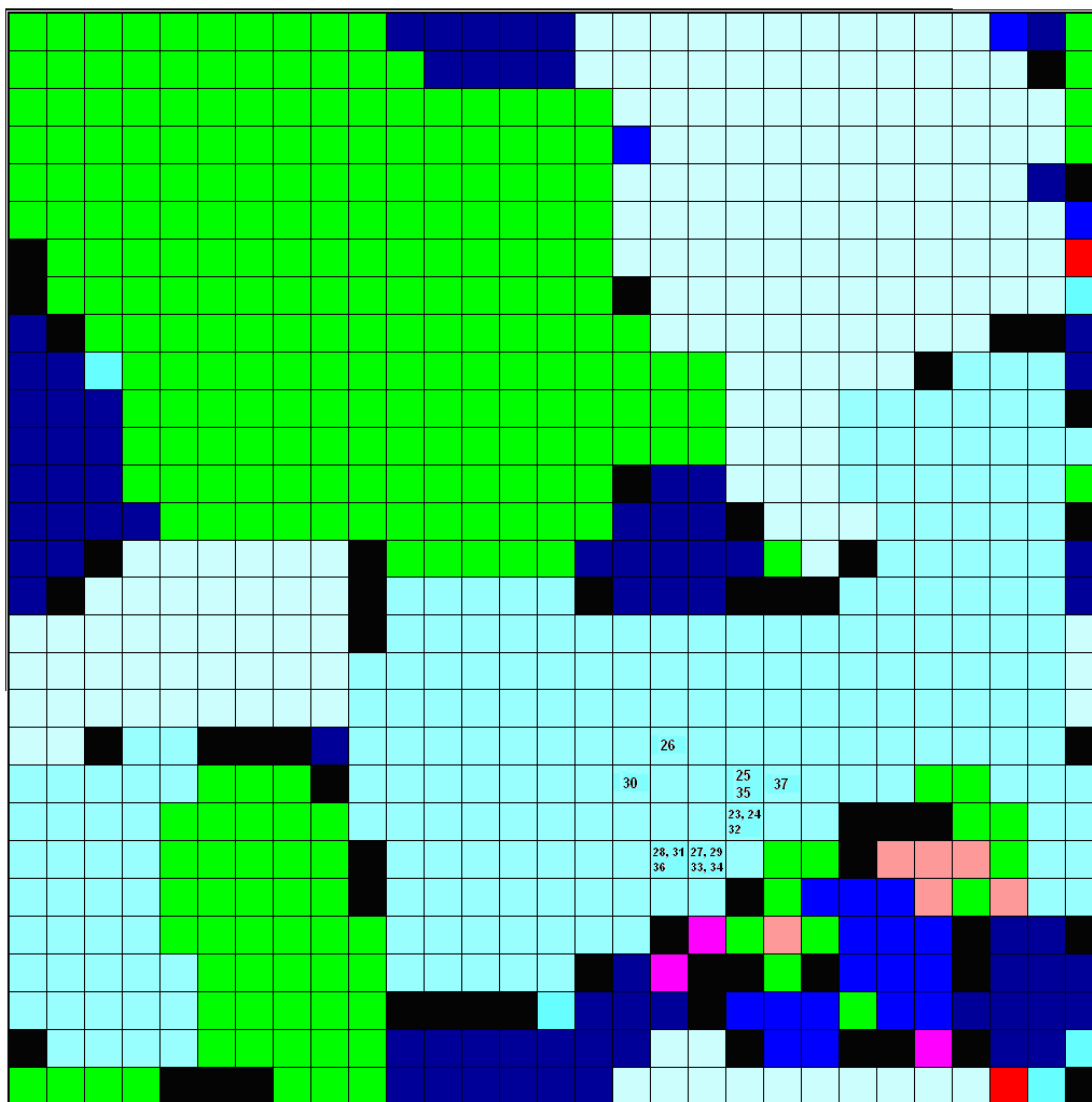


Figura III.17: Mapeamento de 15 carbonilos ($C=O$) em SOM 29×29 baseado em 59 propriedades físico-químicas/topológicas das ligações. Cor verde representa maioritariamente: ligações CC simples, duplas e aromáticas, CN simples e aromáticas, CP simples, CS simples, CH, CB simples, C-Sn simples e CS dupla. Cor azul pálida representa ligações NC simples, dupla, tripla e aromática, NO simples e NN dupla e aromática. Cor azul celeste: OC simples, dupla e aromática, OSi simples e OH simples. Cor azul: ligações ClC, BrC, IC e BC. Cor azul escura: ligações SC simples, dupla e aromática, SN simples, SP simples e dupla, SH simples e SS simples. Cor rosa: ligações FC.

Para o conjunto de 59 descritores Topológicos e Físico-Químicos foi construído mapa auto-organizativo 29×29 que foi utilizado para mapear 15 carbonilos provenientes de 15 moléculas distintas (**Figura III.17** e **Figura III.18**):

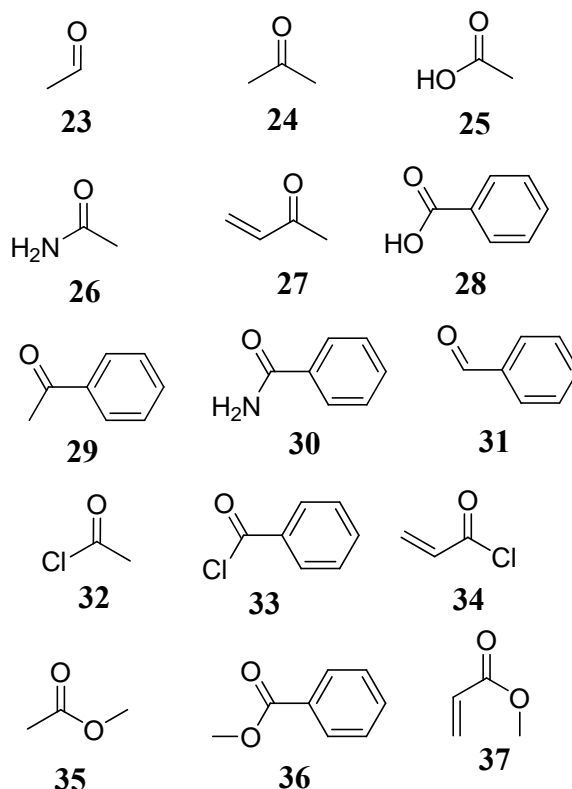


Figura III.18: Compostos contendo ligação carbonilo (C=O) mapeados na Figura III.17.

Analisando a **Figura III.17** tendo em conta a **Figura III.18** pode observar-se que carbonilos de amidas alifáticas **26** e aromáticas **30** estão isoladas em relação aos restantes carbonilos, assim como o carbonilo de éster α,β insaturado **37**. Os carbonilos de ácido carboxílico alifático **25** e éster alifático **35** estão mapeados no mesmo neurónio, ou seja, este mapa não consegue distinguir carbonilos de ésteres e ácidos carboxílicos. Noutro neurónio encontram-se mapeados os carbonilos de aldeído alifático **23**, cetona alifática **24** e cloreto de ácido alifático **32**. Continuando a percorrer a **Figura III.17** encontra-se noutro neurónio os carbonilos de ácido carboxílico aromático **28**, éster aromático **36** e aldeído aromático **31**. Finalmente, noutro neurónio ocupado encontram-se os carbonilos de cetona α,β insaturada **27**, cetona aromática **29**, cloreto de ácido aromático **33** e cloreto de ácido α,β insaturado **34**. Assim sendo, ainda que o ambiente químico influencie na distinção dos diversos carbonilos, a discriminação é apenas parcial. Relativamente ao mapa auto-organizativo representado na **Figura III.17** para além de **a)** carbonilos foram mapeados **b)** ligações carbono-carbono simples, duplas, triplas e aromáticas, **c)** ligações carbono-X, em que X pode ser halogénio ou grupo hidroxilo, **d)** ligações carbono-azoto de amina, imina e nitrilo (avaliação da capacidade de distinguir ordem de ligação), **e)** ligações carbono-azoto de grupos amina e nitro alifáticos e aromáticos (avaliação da capacidade de distinguir estados de oxidação), **f)** ligações carbono-oxigénio de hidroxilos e carbonilos, **g)** ligações carbono-azoto em aminas e amidas (apesar de estado de oxidação do azoto tal como a ordem de ligação ser igual em aminas e em amidas, avalia-se a capacidade de distinguir

funcionalidades), **h)** ligações carbono-carbono aromáticas cujos carbonos podem estar não-substituídos ou substituídos por grupos electro-atractores e/ou doadores.

Com os conjuntos de 45 descritores de ligação topológicos e os 27 descritores físico-químicos foram construídos os respectivos mapas auto-organizativos 29×29 (de forma análoga ao da **Figura III.17**) onde foram mapeados os conjuntos de ligações **a)-h)** tal como foi feito para o caso do conjunto de 59 descritores topológicos/físico-químicos. Os resultados finais aparecem condensados nas **Tabelas III.10, III.11 e III.12**:

Tabela III.10: Capacidade do mapa auto-organizativo produzido com conjunto de 59 descritores topológicos/físico-químicos para discriminar ligações químicas iguais ou semelhantes mas com ambiente químico diferente.

Tipos de Ligação	Em neurónios isolados	No mesmo neurónio
a) carbonilos	<ul style="list-style-type: none"> - Amida alifática - Amida aromática - Éster α,β insaturado 	<ul style="list-style-type: none"> - Ácido carboxílico e éster alifáticos - Aldeído, Cetona e Cloreto de ácido alifáticos - Ácido Carboxílico, Éster e Aldeído aromáticos - Cetona α,β insaturada, Cetona aromática, Cloreto de ácido α,β insaturado, Cloreto de Ácido Aromático
b) ligações carbono-carbono, simples, duplas, triplas, aromáticas	<ul style="list-style-type: none"> - Ligação CC simples -Ligação CC dupla -Ligação CC tripla -Ligação CC aromática 	-
c) ligações carbono-X, em que X é halogénio ou hidroxilo	<ul style="list-style-type: none"> - Ligações CBr - Ligações CHidroxilo 	- Ligações CF e CCl
d) ligações carbono-azoto de amina, imina e nitrilo	<ul style="list-style-type: none"> - Amina - Imina -Nitrilo 	-
e) ligações carbono-azoto de grupos amina e nitro	<ul style="list-style-type: none"> - Amina - Nitro 	-
f) ligações carbono-oxigénio de hidroxilos e carbonilos	<ul style="list-style-type: none"> - Hidroxilos - Carbonilos 	-
g) ligações carbono-azoto em aminas e amidas	<ul style="list-style-type: none"> - Amina -Amida 	-
h) ligações carbono-carbono aromáticas	<ul style="list-style-type: none"> - C=C - C=C-Cl - C=C-NO₂ 	-

Ligações discriminadas aparecem mapeadas em neurónios isolados, caso contrário aparecem mapeadas juntamente com outras ligações distintas no mesmo neurónio.

Tendo em conta os resultados apresentados na **Tabela III.10** é possível constatar que o mapa auto-organizativo consegue separar uma grande porção de ligações iguais ou semelhantes mas com ambiente químico diferente, concretamente, consegue separar ligações CC simples, duplas, triplas e aromáticas, separa igualmente aminas, nitrilos e iminas, tal como ligações CN de aminas e nitros. O mapa proporciona ainda a separação entre aminas e amidas, proporcionando ainda a separação entre

ligações CC aromáticas não substituídas e substituídas por grupos electro-atractores e electro-doadores. Relativamente a carbonilos aparecem inequivocamente separados carbonilos de amidas alifáticas, amidas aromáticas e éster α,β insaturado.

Tabela III.11: Capacidade do mapa auto-organizativo produzido com conjunto de 27 descritores físico-químicos para discriminar ligações químicas.

Tipos de Ligação	Em neurónios isolados	No mesmo neurónio
a) Carbonilos	<ul style="list-style-type: none"> - Amida alifática - Amida aromática 	<ul style="list-style-type: none"> - Ácido Carboxílico Aromático, Ester Aromático e Éster α,β insaturado - Ácido Carboxílico Alifático, Cetona Aromática e Aldeído Aromático - Aldeído Alifático e Cetona Alifática - Cetona α,β insaturada e cloreto de ácido alifático - Cloreto de Ácido α,β insaturado e Cloreto de Ácido Aromático
b) ligações carbono-carbono, simples, duplas, triplas, aromáticas	<ul style="list-style-type: none"> - Ligação CC simples -Ligação CC dupla -Ligação CC tripla -Ligação CC aromática 	-
c) ligações carbono-X, em que X é halogénio ou hidroxilo	<ul style="list-style-type: none"> - Ligações CF - Ligações CHidroxilo 	- Ligações CCl e CBr
d) ligações carbono-azoto de amina, imina e nitrilo	<ul style="list-style-type: none"> - Amina - Imina -Nitrilo 	-
e) ligações carbono-azoto de grupos amina e nitro	<ul style="list-style-type: none"> - Amina -Nitro 	-
f) ligações carbono-oxigénio de hidroxilos e carbonilos	<ul style="list-style-type: none"> - Hidroxilos - Carbonilos 	-
g) ligações carbono-azoto em aminas e amidas	<ul style="list-style-type: none"> - Amina - Amida 	-
h) ligações carbono-carbono aromáticas	-C=C-NO ₂	- C=C-Cl e C=C

Ligações discriminadas aparecem mapeadas em neurónios isolados, caso contrário aparecem mapeadas juntamente com outras ligações distintas no mesmo neurónio.

Observando a **Tabela III.11** é possível constatar que o mapa construído a partir de 27 propriedades físico-químicas de ligações, à semelhança do mapa construído a partir de 59 descritores, permite a separação de ligações CC simples, duplas, triplas e aromáticas, separa igualmente aminas, nitrilos e iminas, tal como ligações CN de aminas e nitros. O mapa proporciona ainda a separação

entre aminas e amidas. Este conjunto de 27 propriedades, distingue-se do caso anterior, por não permitir a separação para ligações CC aromáticas em que os átomos de carbono aparecem substituídos por cloro ou não-substituídos. Este mapa não permite a separação entre ligações carbono-bromo e carbono-cloro, permitindo apenas a separação inequívoca de carbonilos de amidas aromáticas e alifáticas.

Tabela III.12: Capacidade do mapa auto-organizativo produzido com conjunto de 45 descritores topológicos para discriminar ligações químicas.

Tipos de ligação	Em neurónios isolados	No mesmo neurónio
a) carbonilos	<ul style="list-style-type: none"> - Amida Aromática - Aldeído Aromático - Amida Alifática - Cetona Alifática 	<ul style="list-style-type: none"> - Cloreto de Ácido Aromático e Cloreto de Ácido α,β insaturado - Ácido Carboxílico Aromático, Éster Aromático, Éster α,β insaturado - Aldeído alifático e Cloreto de Ácido alifático - Éster Alifático e Ácido Carboxílico Alifático
b) ligações carbono-carbono, simples, duplas, triplas, aromáticas	<ul style="list-style-type: none"> - Ligação CC simples - Ligação CC dupla - Ligação CC tripla - Ligação CC aromática 	-
c) ligações carbono-X, em que X é halogénio ou hidroxilo	-	<ul style="list-style-type: none"> - Ligações CF - Ligações CCl - Ligações CBr (separa Alifáticos de Aromáticos)
d) ligações carbono-azoto de amina, imina e nitrilo	<ul style="list-style-type: none"> - Amina - Imina - Nitrilo 	-
e) ligações carbono-azoto de grupos amina e nitro	-	Ligação CN de amina e nitro
f) ligações carbono-oxigénio de hidroxilos e carbonilos	<ul style="list-style-type: none"> - Hidroxilos - Carbonilos 	-
g) ligações carbono-azoto em aminas e amidas	<ul style="list-style-type: none"> - Amina - Amida 	-
h) ligações carbono-carbono aromáticas	<ul style="list-style-type: none"> - C=C-Cl - C=C - C=C-NO₂ 	-

Ligações discriminadas aparecem mapeadas em neurónios isolados, caso contrário aparecem mapeadas juntamente com outras ligações distintas no mesmo neurónio.

Pela observação à **Tabela III.12** é possível constatar que as 45 características topológicas (ver Metodologia Computacional) não conseguem distinguir uma parcela significativa de ambientes químicos. Concretamente, não foi possível distinguir entre ligações carbono-flúor, carbono-cloro e carbono-bromo, tal como não foram distinguidas ligações carbono-azoto de aminas e de

funcionalidade nitro. Relativamente a carbonilos, foi possível distinguir inequivocamente carbonilos de amida alifática, amida aromática, aldeído aromático e cetona alifática.

Com base nestes resultados antevia-se que os MOLMAPs gerados a partir de 59 descritores topológicos/físico-químicos fossem mais adequados para codificar padrões estruturais que distinguíssem os seis mecanismos reaccionais em causa neste estudo.

Cada um destes três conjuntos de descritores seleccionados foi usado para representar cada uma das 5870 ligações químicas que foram usadas para treinar mapas auto-organizativos (redes de *Kohonen*) de diferentes dimensões (10×10, 15×15, 20×20, 25×25 e 29×29). Os SOMs obtidos para cada conjunto de descritores de ligação foram usados na obtenção de MOLMAPs para representar cada uma das moléculas do conjunto de treino, conjunto de validação e conjunto de previsão. Posteriormente, foram construídos modelos *Random Forest* (1000 árvores e *mtry* = 25) que foram comparados na capacidade de previsão usando conjunto de validação. Os resultados estão na **Tabela III.13**.

Tabela III.13: Comparação entre os diversos modelos Random Forest treinados para prever o mecanismo a partir do substrato representado com MOLMAPs de diversas dimensões baseados em 45 propriedades topológicas de ligação química.

Tipo de MOLMAP (Neurónio vencedor valor 10, neurónios vizinhos valor 3)	Concordância <i>Out of Bag</i> (conjunto de Treino)	Concordância conjunto de validação de 49 compostos
10×10	0,841	0,776
15×15	0,841	0,776
20×20	0,800	0,755
25×25	0,841	0,776
29×29	0,862	0,776

Concordância – Ver Secção II-2.4 b) – Metodologia Computacional. Conjunto de treino e validação.

Tendo em conta os resultados apresentados na **Tabela III.13** para modelos *Random Forest* baseados em 45 propriedades topológicas de ligações químicas pode-se observar variação praticamente negligenciável na capacidade de previsão para conjunto de validação (concordância = 0,755 – 0,776), relativamente aos resultados obtidos a partir do conjunto de treino o valor de concordância é máximo para todos os modelos construídos (concordância = 1,000). Relativamente aos resultados *Out of Bag* (ver Capítulo I) obtidos com o conjunto de treino, observa-se uma variação moderada relativamente à concordância (0,800-0,862). Seguindo o mesmo procedimento foram

construídos modelos *Random Forest* baseados em 27 propriedades físico-químicas de ligação, os resultados foram os seguintes:

Tabela III.14: Comparação, relativamente à concordância, entre os diversos modelos *Random Forest* treinados para prever o mecanismo a partir do substrato representado com MOLMAPs de diversas dimensões baseados em 27 propriedades físico-químicas de ligação química.

Tipo de MOLMAP (Neurónio vencedor valor 10, neurónios vizinhos valor 3)	Concordância <i>Out of Bag</i> (conjunto de treino)	Concordância conjunto de validação de 49 compostos
10×10	0,786	0,735
15×15	0,841	0,714
20×20	0,821	0,755
25×25	0,841	0,755
29×29	0,848	0,755

Conjunto de treino e validação.

Comparando os resultados da **Tabela III.14** com os **Tabela III.13** pode afirmar-se que a qualidade dos resultados é praticamente idêntica. Os resultados da **Tabela III.14** evidenciam uma variação pouco pronunciada da concordância para o conjunto de validação (concordância = 0,714-0,755). Relativamente aos resultados *Out of Bag* obtidos com conjunto de treino, observa-se uma variação mais significativa (concordância = 0,786-0,848). Tendo em conta o conjunto de 59 propriedades físico-químicas das ligações foram obtidos os seguintes resultados (**Tabela III.15**):

Tabela III.15: Comparação, relativamente à Concordância, entre os diversos modelos *Random Forest* treinados para prever o mecanismo a partir do substrato representado com MOLMAPs de diversas dimensões baseados em 59 propriedades físico-químicas/topológicas de ligação química.

Tipo de MOLMAP (Neurónio vencedor valor 10, neurónios vizinhos valor 3)	Concordância <i>Out of Bag</i> (conjunto de treino)	Concordância conjunto de validação de 49 compostos
10×10	0,841	0,735
15×15	0,869	0,776
20×20	0,848	0,816
25×25	0,883	0,837
29×29	0,869	0,796

Conjunto de treino e validação.

Tendo em conta os resultados apresentados na **Tabela III.15** é possível constatar que, tanto MOLMAPs 20×20 como 25×25, ao contrário dos anteriores modelos baseados em 27 e 45 descritores de ligação, permitem ultrapassar valor 0,8 de concordância, com MOLMAPs 25×25 a permitirem atingir o melhor resultado (concordância = 0,837). Variação do valor de *mtry* da *Random Forest* entre 10 e 100 praticamente não influenciou os resultados. Com *mtry* = 10 obtiveram-se previsões para o conjunto independente de teste de 45 moléculas, cujo mecanismo reaccional se baseou em trabalhos publicados por Aptula¹ e Roberts.^{Erro! Marcador não definido.} Os resultados completos deste modelo aparecem condensados na **Tabela III.16**.

Tabela III.16: Resultados obtidos para a classificação de compostos de acordo com o mecanismo reaccional (a-f) nos conjuntos de treino, de validação e de teste.

		<i>Out of Bag</i> (conjunto de treino)	Conjunto de validação de 49 compostos	Conjunto de teste de 45 compostos
Concordância		0,883	0,857	0,756
k		0,847	0,811	0,705
λ_B		0,813	0,839	0,848
Predictibilidade	a	0,774	0,800	0,550
	n° objectos	26	9	12
	b	0,800	0,875	1,000
	n° objectos	19	6	8
	c	0,930	0,857	0,833
	n° objectos	55	19	10
	d	0,895	0,857	1,000
	n° objectos	18	6	11
	e	1,000	0,000	0,000
	n° objectos	8	2	5
Sensibilidade	f	1,000	1,000	-
	n° objectos	19	7	-
	a	0,923	0,889	0,917
	n° objectos	26	9	12
	b	0,842	1,000	0,500
	n° objectos	19	6	8
	c	0,981	1,000	1,000
	n° objectos	55	19	10
	d	0,894	1,000	0,909
	n° objectos	18	6	11
	e	0,875	0,000	0,000
	n° objectos	8	2	5
	f	0,579	0,429	-
	n° objectos	19	7	-
N° objectos total		145	49	45

Definições de concordância, k, λ_B , predictibilidade e sensibilidade aparecem descritas na secção II-2.4 b) – Metodologia Computacional.

Nas estimativas por *Out of Bag* com o conjunto de treino, as percentagens de acertos são elevadas, excepto a sensibilidade relativamente à classe “f” (Outros mecanismos reaccionais) que é

significativamente mais baixa (0,579). Esta classe envolve variedade de mecanismos e dado o reduzido tamanho do conjunto de treino é fácil haver novos compostos cujo mecanismo reaccional não aparece abrangido no conjunto de treino. Para o conjunto de validação de 49 compostos a sensibilidade foi baixa para a classe “f” (0,429) e tanto a predictibilidade como a sensibilidade relativamente à classe “e” (adição de Michael) são nulas. Isto acontece já que os oito compostos do conjunto de treino pertencentes à classe “e” são exclusivamente sulfonas e sulfonatos α,β insaturados. Por sua vez os objectos do conjunto de teste pertencentes à classe “e” são compostos carbonílicos α,β -insaturados. Por esse motivo as previsões relativamente a essa classe foram incorrectas.

Para este conjunto as restantes predictibilidades e sensibilidades são elevadas, tal como as percentagens de acertos que também são elevadas (concordância = 0,857). Finalmente este modelo foi testado com conjunto de teste de 45 compostos observando-se uma concordância de 0,756, mas predictibilidade e sensibilidade nulas para a classe “e”. Adicionalmente observa-se predictibilidade baixa para classe “a” (0,550) tal como sensibilidade baixa para classe “b” (0,500). As restantes sensibilidade e predictibilidades são elevadas. Os resultados de sensibilidades, predictibilidades, concordância, k e λ_B baseiam-se em Tabela de contingência (**Tabela III.17**).

Tabela III.17: Tabela de contingência relativamente aos resultados obtidos com o conjunto de teste de 45 moléculas.

		Experimental					Total linha
		a	b	c	d	e	
Previsão	a	11	4	0	0	5	20
	b	0	4	0	0	0	4
	c	1	0	10	1	0	12
	d	0	0	0	10	0	10
	e	0	0	0	0	0	0
Total coluna		12	8	10	11	5	

Na diagonal principal aparecem representados aqueles objectos em que há correspondência entre as classes reaccionais (a-e) experimentais e previstas.

Estes resultados podem ser explicados por análise detalhada a cada um dos compostos, presentes neste conjunto de 45 moléculas, cujas previsões não coincidem com a classe de mecanismo reaccional expectável,^{1,26} de acordo com **Tabela III.18**. Para a interpretação das previsões recorreu-se também à inspecção dos compostos no conjunto de treino a que a RF atribui semelhança mais elevada como composto a prever. A molécula **49**, um éster, é prevista como sendo classe “a” (formação de base de Schiff) quando na realidade reage por mecanismo “b” associado a grupos acilo. A razão para esta previsão incorrecta deve-se à baixa proximidade da molécula **49** a objectos do conjunto de treino

(proximidade máxima = 0,281). A probabilidade desta previsão (“a”) é igualmente muito baixa (0,334) muito próxima da probabilidade de reagir segundo a classe (“b”) correcta (0,316).

Relativamente às moléculas **32** e **50** também são incorrectamente associadas a classe “a” por modelo RF quando na realidade são cloretos de ácido que reagem segundo mecanismo “b”. Nestes casos foi escolhido mecanismo “a” já que os carbonilos de cloreto de ácido alifáticos vão activar alguns dos mesmos neurónios activados por carbonilos de aldeído alifático (reage segundo mecanismo “a”) em mapa auto-organizativo onde são mapeadas as ligações. Ambas as funcionalidades vão corresponder a algumas posições em comum em MOLMAP, desta forma é mais difícil ao modelo RF distinguir entre dois tipos de moléculas, principalmente se forem escassas as ligações que as distingam. Um caso relativamente semelhante é o da molécula **33**, um cloreto de ácido aromático que é associado por modelo RF à classe “a” quando na realidade reagiria por mecanismo “b”, os carbonilos de cloreto de ácido aromático activam alguns dos mesmos neurónios activados por carbonilos de cetona aromática (reage segundo mecanismo “a”) em mapa auto-organizativo. Assim, há troca no mecanismo proposto pelo modelo RF.

Relativamente às moléculas **27**, **51** e **52**, cetonas α,β -insaturadas, reagem segundo mecanismo “e” (adição de Michael) mas o modelo RF prevê que estes casos vão reagir segundo mecanismo “a” (formação de base de Schiff), sendo as moléculas mais próximas no conjunto de treino cetonas alifáticas. Isto deve-se ao facto que no conjunto de treino existirem poucas moléculas que reagem segundo mecanismo de adição de Michael e nenhuma das moléculas contém funcionalidade carbonílica α,β -insaturada (são sulfonas e sulfonatos α,β -insaturados). Adicionalmente estas moléculas são relativamente complexas contendo outras funcionalidades. E como estão a ser considerados todos os fragmentos funcionalizados de cada molécula e não exclusivamente o centro reaccional, é natural que moléculas simples como o caso de **27**, **51** e **52** possam não ter como mais próximas moléculas que reajam segundo mecanismo “e”.

Uma análise semelhante pode ser feita em relação às moléculas **53** e **54**, aldeídos α,β insaturados que apesar de se esperar que reajam segundo mecanismo “e” na realidade o modelo RF prevê que estas duas moléculas vão reagir segundo mecanismo “a”, isto já que a molécula mais próxima no conjunto de treino é o formaldeído, molécula simples contendo funcionalidade carbonílica, à semelhança do que acontece com as moléculas **53** e **54**. A razão para que outras moléculas do conjunto de treino a reagir segundo mecanismo “e” não tenham sido seleccionadas como as mais próximas é a mesma que para os casos das moléculas **27**, **51** e **52**.

Tabela III.18-1: Compostos pertencentes ao conjunto de teste de 45 moléculas e respectivas classes de mecanismo reaccional, de acordo com a literatura e classe prevista por modelo RF.

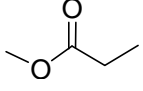
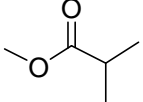
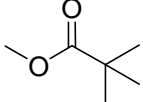
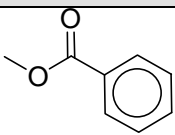
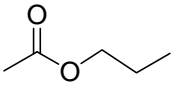

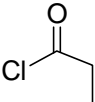
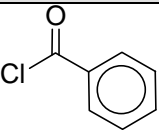
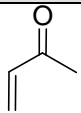
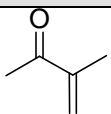
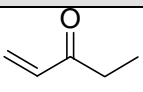
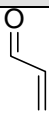
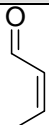

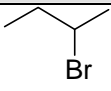
		
b (lit)	b (lit)	b (lit)
b (prev)	b (prev)	b (prev)
46	47	48
		
b (lit)	b (lit)	b (lit)
b (prev)	a (prev)	a (prev)
36	49	32
		
b (lit)	b (lit)	e (lit)
a (prev)	a (prev)	a (prev)
50	33	27
		
e (lit)	e (lit)	e (lit)
a (prev)	a (prev)	a (prev)
51	52	53
		
e (lit)	c (lit)	c (lit)
a (prev)	c (prev)	c (prev)
54	55	56

Tabela III.18-2: Continuação da apresentação dos compostos pertencentes ao conjunto de teste de 45 moléculas e respectivas classe de mecanismo reaccional, de acordo com a literatura, e classe prevista por modelo RF.

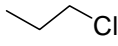
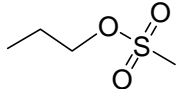

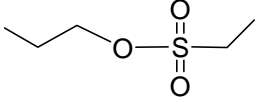
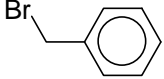

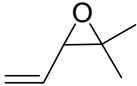
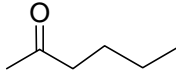
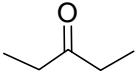

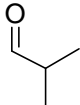
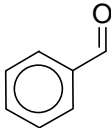
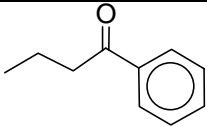
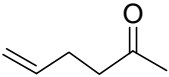
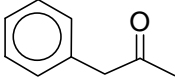
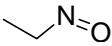
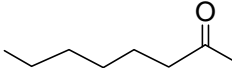
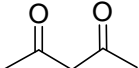
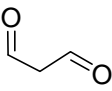
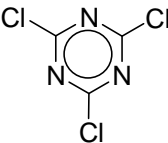
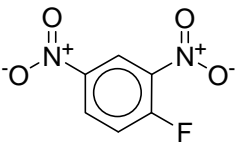
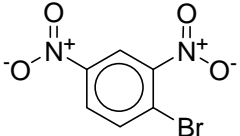
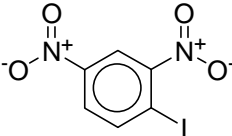
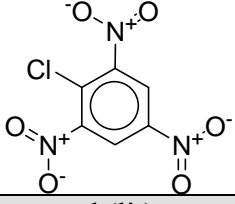
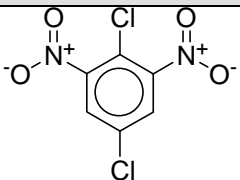
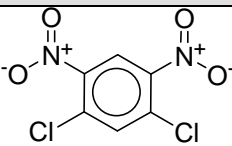
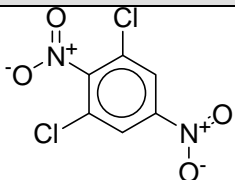
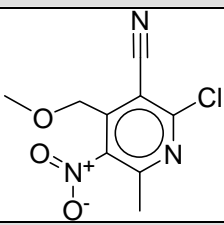
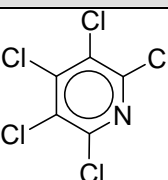
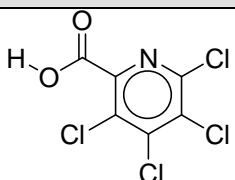
		
c (lit)	c (lit)	c (lit)
c (prev)	c (prev)	c (prev)
57	58	59
		
c (lit)	c (lit)	c (lit)
c (prev)	c (prev)	c (prev)
60	61	62
		
c (lit)	a (lit)	a (lit)
c (prev)	a (prev)	a (prev)
63	64	65
		
a (lit)	a (lit)	a (lit)
a (prev)	a (prev)	a (prev)
66	67	31
		
a (lit)	a (lit)	a (lit)
a (prev)	a (prev)	a (prev)
68	69	70

Tabela III.18-3: Continuação da apresentação dos compostos pertencentes ao conjunto de teste de 45 moléculas e respectivas classe de mecanismo reaccional, de acordo com a literatura, e classe prevista por modelo RF.

		
a (lit)	a (lit)	a (lit)
c (prev)	a (prev)	a (prev)
71	72	73
		
a (lit)	d (lit)	d (lit)
a (prev)	c (prev)	d (prev)
74	75	76
		
d (lit)	d (lit)	d (lit)
d (prev)	d (prev)	d (prev)
77	78	79
		
d (lit)	d (lit)	d (lit)
d (prev)	d (prev)	d (prev)
80	81	82
		
d (lit)	d (lit)	d (lit)
d (prev)	d (prev)	d (prev)
83	84	85

Continuando a percorrer a **Tabela III.18**, a seguinte molécula mal classificada é a molécula **71** que, reagindo segundo mecanismo de formação de base de Shiff (“a”), é prevista pelo modelo como reagindo segundo mecanismo “c” (S_2N). Esta discrepância deve-se ao facto desta molécula ser pouco semelhante a moléculas do conjunto de treino (a proximidade máxima é 0,359). Por sua vez a probabilidade de pertença a qualquer classe reaccional também é baixa. Relativamente à classe “c” (a classe mais provável segundo o modelo) a probabilidade é de apenas 0,424, seguida de perto pela probabilidade de pertença à classe “a” (0,308), a segunda classe mais provável e na realidade a classe acertada. Finalmente, a molécula **75**, que, reagindo segundo mecanismo “d” de substituição nucleofílica aromática (S_NAr), vai ter como previsão do modelo o mecanismo “c” (S_2N). Isto deve-se ao facto da molécula mais próxima no conjunto de treino ter funcionalidades que podem reagir segundo mecanismos “c” ou “d”, reagindo selectivamente na presença de nucleófilo por mecanismo “c”.

Tendo em conta estes resultados o modelo de classificação construído apresenta limitações derivadas de alguma incapacidade de discriminação por parte dos mapas auto-organizativos onde são mapeadas as ligações. Também a estratégia de não considerar apenas ligações do centro reaccional (de forma a generalizar este método a bases de dados sem centro reaccional identificado) tem a desvantagem de poderem ser consideradas relevantes tipos de ligação que na realidade não o são, e esse problema surgiu em casos pontuais.

De forma a saber quais as funcionalidades (posições no MOLMAP) preponderantes na classificação segundo mecanismo reaccional, foram analisados os descritores mais importantes para o modelo RF (MOLMAPs 25×25, 1000 árvores, $mtry = 10$). Foi também construída uma Árvore de Decisão como suporte a essa interpretação dado ser um método de análise simples e ao mesmo tempo o algoritmo usado ser semelhante a *Random Forest*. A Árvore de Decisão está representada na **Figura III.19**.

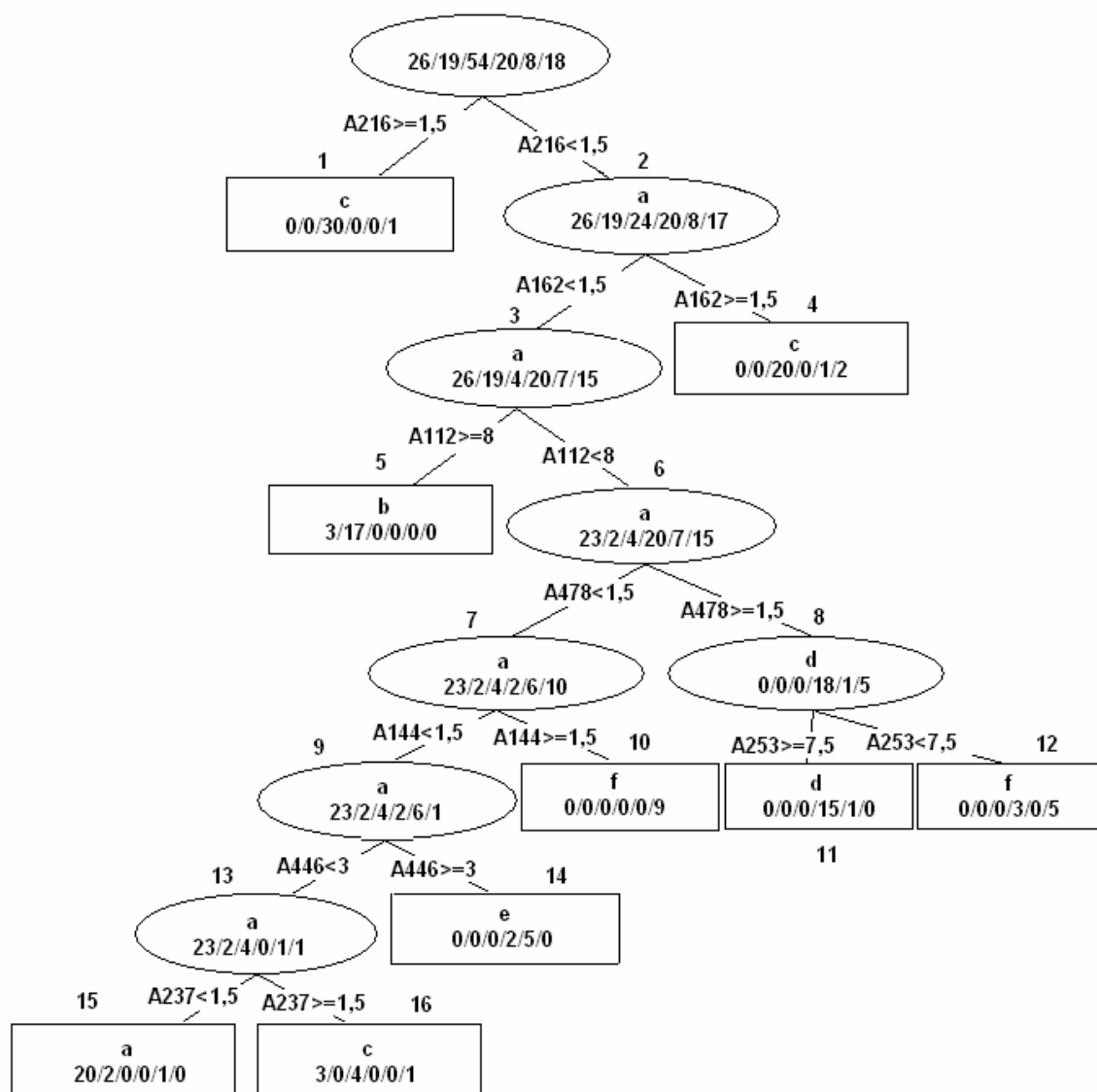


Figura III.19: Árvore de Decisão construída com MOLMAPs 25x25 aplicada ao estudo de classificação de compostos quanto ao mecanismo de reacção na presença de butilamina.

A primeira partição baseia-se no descritor A216, que representa a presença ligações carbono-halogénio (cloro, bromo e iodo), sendo possível isolar de uma forma quase total 30 objectos da classe “c” (substituição nucleofílica) no nó terminal 1. O outro subconjunto, 2, vai sofrer partição, a partir do descritor A162 que assinala a presença de ligações carbono-oxigénio simples características de epóxidos, resultando em dois nós, 3 e 4. Este último, associado à presença de ligação C-O, agrupa 20 compostos de classe “c” predominantemente epóxidos. O nó 3 vai sofrer partição, a partir de descritor A112, que corresponde preponderantemente a ligação carbono-oxigénio simples de ésteres, e que isola uma quantidade significativa (17) de compostos de classe “b” (acilações), maioritariamente ésteres. O nó 6, por sua vez, vai sofrer partição a partir do descritor A478 que representa ligação

carbono-cloro predominantemente em anel aromático, e que permite isolar maioritariamente compostos que reagem através de mecanismo “d” (substituição nucleofílica aromática). Este nó (8) vai ser ainda purificado com o descritor A253, que assinala a presença de ligações carbono-carbono aromáticas. Por sua vez o nó 7 sofre partição pelo descritor A144, que representa ligações duplas carbono-enxofre, associadas a isotiocianatos e compostos tiocarbonílicos que reagem através de outros mecanismos (classe “f”).

Com a partição do descritor A144 vai resultar igualmente o nó 9, que vai sofrer partição através de descritor A446, que assinala maioritariamente a presença de ligações duplas enxofre-oxigénio de sulfonas, que no conjunto de treino aparecem conjugadas a olefina. Desta forma foi possível obter o nó 14 (presença de sulfona α,β insaturada), isolando maioritariamente objectos de classe “e” (adição de Michael). O outro nó resultante (13) vai sofrer partição a partir do descritor A237, associado a ligações carbono-enxofre simples mas não só, resultando os nós terminais 15 e 16. No nó terminal 16 predominam objectos de classe “a” e “c”, constituído por uma grande quantidade de tioepóxidos (mecanismo “c”), aldeídos e cetonas (mecanismo “a”), estes aldeídos e cetonas incluem ligações carbono-enxofre sendo esse o motivo para que estejam colocados neste nó terminal. O modelo Árvore de Decisão permite obtenção de concordância moderadamente elevada (0,714) para conjunto de validação de 49 compostos (0,855 para conjunto de treino).

Relativamente ao modelo RF os descritores mais importantes são os seguintes (**Tabela III.19**):

Tabela III.19: Os 10 descritores mais importantes seleccionados por modelo *Random Forest* e respectiva medida de importância.

Posição	Descritor	Importância (ver Capítulo I)
1	A191	1,328
2	A240	1,231
3	A215	1,183
4	A521	1,182
5	A216	1,144
6	A241	1,097
7	A575	1,074
8	A383	1,063
9	A190	1,055
10	A384	1,043

Tendo em conta a lista dos dez descritores mais importantes segundo modelo RF, é possível observar que uma quantidade considerável destes descritores representa ligações carbono-halogénio, tal como o descritor A191 que aparece na 1ª posição e que representa ligações carbono-bromo com

bromo em posição benzílica. Por sua vez na 2ª posição surge descritor A240, associado a ligação carbono-bromo alifática. Na 3ª posição surge descritor A215 associado a ligação carbono-bromo com bromo em posição alílica. Na 5ª posição aparece descritor A216 representando ligações carbono-cloro, carbono-bromo e carbono-iodo alifáticas.

Relativamente à 6ª posição, esta é ocupada por descritor A241, associado a ligações carbono-bromo e carbono-cloro alifáticas, alílicas e benzílicas. Já na 9ª posição surge descritor A190 que representa ligações carbono-oxigénio de epóxidos, ligações carbono-cloro alifáticas e ligações carbono-bromo olefinicas. Todos estes descritores estão relacionados com o mecanismo “c” de substituição nucleofílica.

Na 4ª posição surge descritor A521 que representa carbonilos de cetonas alifáticas e aromáticas, representando mecanismo “a” de formação de base de Schiff. Na 7ª posição está situado descritor A575, que representa ligação carbono-oxigénio simples de ésteres e ácidos carboxílicos, este descritor representa mecanismo “b” de reacção de grupo acilo. Na 8ª posição está o descritor A383 a representar ligações carbono-carbono associadas a carbonilos de aldeído e éster. No que diz respeito à 10ª posição esta está representada pelo descritor A384, que representa ligações carbono-carbono de aldeído e ligação carbono-hidrogénio adjacente a funcionalidade éster. Os descritores situados na 8ª e 10ª posições representam em simultâneo mecanismos de reacção “a” e “b”.

De uma forma resumida os dez descritores mais relevantes seleccionados por modelo RF representam mecanismos a, b, c de reacção, o mecanismo “c” é sem dúvida o mais representado tanto por descritores colocados nas dez primeiras posições como por número de objectos no conjunto de treino (ver nó inicial em **Figura III.19**). O mecanismo “a” o segundo mais representado no conjunto de treino está de alguma forma associado a dois descritores colocados nas dez primeiras posições. Também representado por dois descritores nas dez primeiras posições surge o mecanismo “b” que é o quarto mecanismo mais representado no conjunto de treino.

Até aqui foram discutidos modelos baseados nos algoritmos de *Random Forest* e *Árvore de Decisão*, métodos que operam bem sobre uma quantidade inicial de descritores elevada, seguidamente será discutido o uso de redes neuronais de *Kohonen* para a classificação de compostos segundo mecanismo reaccional na presença de nucleófilo. Regra geral, as redes neuronais de *Kohonen* operam sobre um conjunto de descritores reduzido, sendo aconselhável uma selecção prévia de descritores ou alguma forma de os potenciar. Contrariamente aos algoritmos *Random Forest* e *Árvore de Decisão*, as redes neuronais de *Kohonen* operam sobre todos os descritores moleculares em simultâneo. Tendo em conta o objectivo proposto, classificar compostos segundo mecanismo reaccional, e as características associadas às redes neuronais de *Kohonen* foram seguidas três estratégias distintas. Em todas as

estratégias usou-se como ponto de partida MOLMAPs 15×15 do tipo 10_3 normalizados segundo normalização Z, ou seja, inicialmente cada molécula vai estar representada por 225 descritores. Esses descritores inicialmente foram utilizados para treinar um modelo baseado numa rede neuronal de *Kohonen* 13×13 treinada em 100 *epochs*. Este modelo serviu de ponto de comparação com outros três modelos construídos, ora com subconjuntos seleccionados dos 225 descritores MOLMAP iniciais, ora com estes 225 descritores MOLMAP de reagente adicionados de 225 descritores MOLMAP de reacções.

Num caso foram seleccionados 32 descritores MOLMAP através de algoritmos genéticos (ver Capítulo I). Noutro caso a selecção baseou-se na identificação de descritores associados à perda ou alteração de ligação na conversão de reagentes a produtos (descriptor MOLMAP com valor inferior nos produtos do que nos reagentes). Foram assim seleccionados os 68 descritores MOLMAP. Em ambos os casos os descritores seleccionados foram usados para treinar redes de *Kohonen* 13×13 ao longo de 100 *epochs*.

Na terceira estratégia cada composto foi representado por um vector de 450 coordenadas (225 correspondendo a MOLMAP de reagente e 225 a MOLMAP de reacção). Ao treinar uma rede de *Kohonen* com estes objectos, utiliza-se informação implícita acerca do centro reaccional (proporcionada por MOLMAP de reacção), o que não acontece com MOLMAP de reagente. Posteriormente a rede de *Kohonen* assim treinada foi modificada apagando os pesos correspondentes aos 225 descritores MOLMAP de reacção permanecendo os 225 pesos relativos aos 225 descritores MOLMAP de reagente. Desta forma a previsão do mecanismo reaccional para novas moléculas poderá basear-se em exclusivo na molécula em estudo não sendo necessário saber *à priori* o mecanismo reaccional dessa molécula na presença de nucleófilo butilamina. Para a construção deste modelo foi utilizada uma rede de *Kohonen* 13×13 treinada em 100 *epochs*. Os modelos construídos, todos baseados em redes de *Kohonen*, foram comparados na capacidade de previsão, tanto para o conjunto de treino de 145 moléculas, como para o conjunto de validação de 49 moléculas – **Tabela III.20.**

Tabela III.20: Comparação de quatro modelos baseados em redes neuronais de *Kohonen* relativamente à capacidade de classificação, de acordo com o mecanismo reaccional (a-f), referente ao conjunto de treino e conjunto independente de validação.

	MOLMAP 25×25 A totalidade dos 225 descritores de reagente (ponto de partida) (a)		MOLMAP 25×25 32 descritores de reagente seleccionados por algoritmos genéticos (b)		MOLMAP 25×25 68 descritores de reagente seleccionados por observação de desaparecimento de ligações em MOLMAP de reacção (c)		MOLMAP 25×25 previsão baseada em 225 descritores de reagentes mas treino de modelo baseou-se em 225 descritores de reagente e 225 descritores de reacção (d)	
	Treino	Val.	Treino	Val.	Treino	Val.	Treino	Val.
Concordância	0,703	0,469	0,924	0,694	0,628	0,367	0,462	0,449
k	0,640	0,333	0,903	0,600	0,556	0,210	0,298	0,259
λ_B	0,522	0,267	0,878	0,533	0,400	0,100	0,344	0,300
Predictibilidade	a	0,947	0,500	0,960	0,750	0,818	0,500	0,606
	nº	26	9	26	9	26	9	26
	b	0,857	0,143	1,000	0,750	0,923	0,000	0,222
	nº	19	6	19	6	19	6	19
	c	0,930	0,706	1,000	0,900	0,921	0,556	0,641
	nº	55	19	55	19	55	19	55
	d	0,941	0,833	0,895	0,571	0,875	0,500	0,200
	nº	18	6	18	6	18	6	18
	e	1,000	0,000	1,000	0,000	1,000	0,000	0,000
	nº	8	2	8	2	8	2	8
Sensibilidade	f	0,917	0,500	0,900	0,500	0,889	0,333	0,000
	nº	19	7	19	7	19	7	19
	a	0,692	0,444	0,923	0,667	0,692	0,333	0,769
	nº	26	9	26	9	26	9	26
	b	0,923	0,250	0,941	0,600	0,857	0,000	0,211
	nº	19	6	19	6	19	6	19
	c	0,727	0,632	0,945	0,947	0,636	0,526	0,745
	nº	55	19	55	19	55	19	55
	d	1,000	0,833	1,000	0,800	1,000	0,800	0,111
	nº	18	6	18	6	18	6	18
Nº objectos	e	1,000	0,000	0,875	0,000	1,000	0,000	0,000
	nº	8	2	8	2	8	2	8
	f	0,578	0,143	0,947	0,429	0,421	0,143	0,000
	nº	19	7	19	7	19	7	19
Nº objectos		145	49	145	49	145	49	145

Definições de concordância, k, λ_B , predictibilidade e sensibilidade aparecem descritas na secção II-2.4 b) – Metodologia Computacional.

É possível observar que os melhores resultados são obtidos por modelo (b), baseado em 32 descritores MOLMAP de reagentes, seleccionados a partir de algoritmos genéticos. Este modelo permitiu a obtenção de valores elevados de concordância para o conjunto de treino (0,924) e

moderadamente elevados para o conjunto de validação (0,694). Os resultados são substancialmente mais elevados que os obtidos pelo modelo (a) que utilizou todos os 225 descritores MOLMAP (concordância = 0,703 para o conjunto de treino e 0,469 para o conjunto de validação). Por sua vez, os modelos (c) e (d) apresentam resultados bastante pobres (concordância = 0,628 e 0,462 respectivamente para o conjunto de treino e 0,367 e 0,449 respectivamente para o conjunto de teste).

Analisando em maior detalhe os resultados apresentados pelo modelo (b) podem-se observar valores elevados de sensibilidade e predictibilidade para conjunto de treino enquanto que para conjunto de validação relativamente à predictibilidade é possível observar valores elevados para as classes “a”, “b” e “c” (0,750, 0,750 e 0,900 respectivamente) enquanto que para as classes “d” e “f” as predictibilidades são relativamente baixas (0,571 e 0,500 respectivamente). Finalmente a predictibilidade para a classe “e” é nula. Para tentar explicar estes resultados importa sublinhar que a classe “e” é a menos representada no conjunto de treino (8 compostos), estando constituída por sulfonas e sulfonatos α,β insaturados, enquanto que os 2 compostos de classe “e” no conjunto de validação contêm funcionalidades não representadas no conjunto de treino no âmbito desta classe. Um dos compostos é amida α,β insaturada, sendo o carbonilo confundido com carbonilo de cetona ou aldeído e levando a previsão de classe “a”. Neste caso os MOLMAP utilizados revelam pouca capacidade de discriminação relativamente a carbonilos contidos em diferentes funcionalidades. O outro objecto de classe experimental “e” é um sulfóxido α,β insaturado, funcionalidade que não se encontra presente no conjunto de treino. Por sua vez as classes “d” e “f” aparecem seguidamente como as classes menos representadas no conjunto de treino com 18 e 19 compostos respectivamente. Em ambos os casos há objectos representados no conjunto de validação cujas funcionalidades não correspondem exactamente às funcionalidades presentes no conjunto de treino. A classe “c” que evidencia predictibilidade mais elevada é claramente a mais representada no conjunto de treino (55 compostos). A classe “c” é aquela que apresenta sensibilidade mais elevada (0,947) seguida pelas classes “a”, “b” e “d” (0,667, 0,600 e 0,800 respectivamente). Finalmente as classes “f” e “e” apresentam as sensibilidades mais baixas (à semelhança da predictibilidade) apresentando como valores 0,429 e 0 respectivamente.

Os restantes erros no conjunto de validação podem ser explicados pela presença de múltiplas funcionalidades potencialmente reactivas (segundo mecanismos diferentes) na mesma molécula, presença de funcionalidades incomuns ou mesmo inexistentes no conjunto de treino e insuficiente capacidade de discriminação entre carbonilos contidos em diferentes funcionalidades, assim como incapacidade de discriminar ligações carbono-cloro olefinicas e aromáticas.

Discutindo ainda o modelo (b) mas do ponto de vista do agrupamento entre classes efectuado por rede de *Kohonen* é possível observar por análise à **Figura III.20** que o modelo baseado em rede

de *Kohonen* agrupa bem as diferentes classes reaccionais do conjunto de treino. As classes “A”, “B”, “C” e “D” representadas a vermelho, azul escuro, verde e amarelo respectivamente estão particularmente bem agrupadas. Nesta análise há que ter em conta que a superfície (mapa) apresentada na **Figura III.20** tem na realidade topologia toroidal – a extremidade superior do mapa é considerada adjacente à extremidade inferior e a extremidade direita à extremidade esquerda. Pode-se assim considerar que estas quatro classes estão bem agrupadas. As classes “E” e “F” (representadas a azul claro e rosa respectivamente) encontram-se parcialmente agrupadas. Por um lado estas duas classes são das menos representadas no conjunto de treino, tendo pouco peso na obtenção do modelo, principalmente a classe “E”. Já a classe “F” encontra-se representada por uma panóplia variada de funcionalidades (e de mecanismos reaccionais também) sendo ainda mais difícil agrupar esta classe. Observam-se ainda neurónios a negro, já que são neurónios onde há conflitos entre diferentes classes não havendo uma classe atribuída.

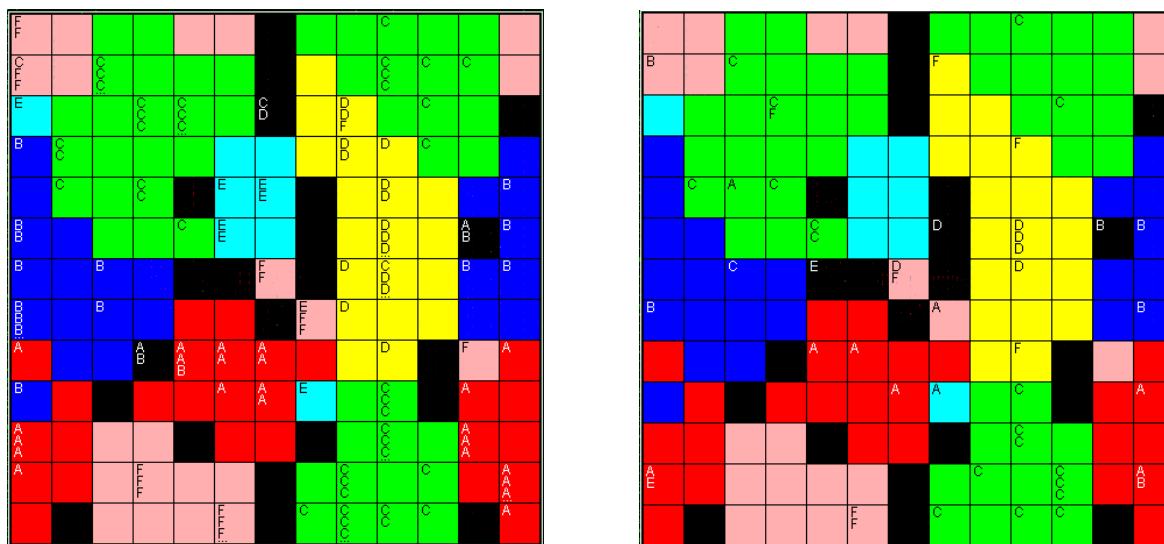


Figura III.20: Mapa de *Kohonen* onde estão mapeados conjunto de treino – lado esquerdo e conjunto de validação – lado direito. Rede de *Kohonen* treinada com 32 descritores MOLMAP de reagente seleccionados por algoritmos genéticos. Vermelho – classe A, Azul escuro – classe B, Verde – classe C, Amarelo – classe D, Azul claro – classe E e Rosa – classe F.

Na **Figura III.20** é possível observar o mesmo mapa representado duas vezes. Num dos mapas (lado esquerdo) estão mapeados objectos do conjunto de treino e no mapa do lado direito estão mapeados objectos do conjunto de validação. Relativamente ao mapeamento do conjunto de treino pode-se observar que a grande maioria dos objectos estão correctamente classificados daí a elevada concordância apresentada na **Tabela III.20** (0,924). Relativamente ao conjunto de validação pode-se observar que uma parcela diminuta mas significativa de objectos se encontra mal classificada daí a moderada concordância observada (0,694). Observa-se ainda que as zonas correspondentes às classes

“A”, “B” e “C” possuem maior proporção de objectos correctamente mapeados, de acordo com as predictibilidades observadas na **Tabela III.20** (0,750, 0,750 e 0,900 respectivamente). Apesar da capacidade evidenciada por modelo baseado em rede de *Kohonen* (**Figura III.20**), modelo *Random Forest* usando MOLMAPs 25×25 (625 descritores) e considerando *mtry* = 10 (**Tabela III.20**) apresenta claramente melhor capacidade de previsão (concordância = 0,857 para conjunto de validação de 49 compostos vs 0,694 em modelo baseado em rede de *Kohonen*). O modelo baseado em Árvore de Decisão apresenta concordância para conjunto de validação igual a 0,714. Apesar de apresentarem pior capacidade de previsão, os modelos baseados em rede de *Kohonen* e Árvore de Decisão são mais interpretáveis que o modelo baseado em *Random Forest*.

III-3.2 b): Estudo do Mecanismo Reaccional de Reacções na Presença de butilamina

Nesta secção estudam-se as mesmas reacções que serviram como base ao estudo de classificação de reagentes, mas os objectos são agora reacções em vez de reagentes. Estas reacções foram classificadas segundo os mesmos tipos de reacções consideradas no estudo **a)**: **i)** formação de bases de Schiff (classe a), **ii)** reacção de acilação (classe b), **iii)** substituição nucleofílica (classe c), **iv)** substituição nucleofílica aromática (S_NAr) (classe d), **v)** adição de Michael (classe e) e **vi)** outros tipos de reacção (classe f), de acordo com **Esquemas III.1-III.6** contidos na Introdução Teórica.

Para a construção de modelos de classificação segundo mecanismo reaccional e, de acordo com Metodologia Computacional, foram seguidos os mesmos passos que foram utilizados para o estudo **a)** anterior: **i)** standardização de reacções, **ii)** selecção de reacções, **iii)** selecção de conjuntos de ligações para treinar Redes de *Kohonen*, **iv)** selecção dos melhores conjuntos de descritores de ligação química, **v)** normalização de descritores de ligações, **vi)** treino de mapas auto-organizativos (SOM) baseados em redes de *Kohonen* de diferentes dimensões, **vii)** obtenção de descritores MOLMAP, **viii)** construção de conjuntos de dados derivados de bases de dados reaccionais e **ix)** construção de modelos de classificação. Relativamente aos tópicos **iv)-vii)** foram usados mapas organizativos treinados a partir de conjuntos de ligações representadas cada uma por 59 propriedades topológicas/físico-químicas, os mesmos mapas e as mesmas propriedades definidos como mais adequados no estudo **a)** anterior. Concretamente foram usados mapas 10×10, 15×15, 20×20, 25×25 e 29×29 que foram utilizados na obtenção de MOLMAPs com as mesmas dimensões, do tipo 10_3. Esses MOLMAP representavam 145 reacções do conjunto de treino, correspondendo aos 145 reagentes usados com conjunto de treino no estudo **a)** anterior. Da mesma forma foram representados por MOLMAPs o conjunto de validação de 49 reacções e o conjunto de teste de 45 reacções, correspondendo respectivamente aos conjuntos de 49 e 45 reagentes usados no estudo **a)**.

O MOLMAP de uma reacção é obtido pela subtracção dos MOLMAPs dos reagentes (reagente e butilamina) ao MOLMAP do produto. Neste caso, um valor negativo numa determinada posição do MOLMAP resulta da alteração ou desaparecimento de uma ligação, enquanto que ligações que se mantêm com propriedades iguais ou semelhantes dão origem a valores nulos, e valores positivos resultam da formação de uma nova ligação (ou do aparecimento de novas propriedades numa ligação). Esta informação será útil quando for necessário interpretar as características estruturais mais relevantes para a classificação de reacções.

Para a construção de modelos de classificação foi usado o algoritmo *Random Forest* (1000 árvores e *mtry* = 25). Estes modelos foram comparados na capacidade de previsão usando conjunto de validação de 49 compostos – **Tabela III.21**.

Tabela III.21: Comparação, relativamente ao valor de concordância, entre os diversos modelos *Random Forest* construídos com MOLMAPs de diversas dimensões baseados em 59 propriedades físico-químicas de ligação química.

Tipo de MOLMAP (Neurónio vencedor valor 10, neurónios vizinhos valor 3)	Concordância conjunto de treino 145 compostos	Concordância <i>Out of Bag</i> (conjunto de treino)	Concordância conjunto de validação de 49 compostos
10×10	0,986	0,910	0,898
15×15	0,993	0,966	0,959
20×20	1,000	0,917	0,980
25×25	1,000	0,931	0,939
29×29	1,000	0,924	0,959

Conjuntos de treino e validação.

É possível constatar que a percentagem de objectos bem classificados em qualquer dos conjuntos considerados é francamente superior às percentagens de acertos evidenciadas no estudo **a)** (comparar com **Tabela III.16**) principalmente no estudo *Out of Bag* e conjunto de validação de 49 compostos. Analisando detalhadamente a **Tabela III.21** é possível ainda observar que MOLMAPs 15×15 são aqueles a permitir uma maior constância relativamente à capacidade de previsão (concordâncias = 0,966 e 0,959 para estudo *Out of Bag* e conjunto de validação respectivamente). Este tipo de MOLMAP apresenta o melhor resultado no estudo *Out of Bag*. Por estas razões, MOLMAPs 15×15 foram usados na optimização do valor de *mtry* relativo à *Random Forest* (**Tabela III.22**):

Tabela III.22: Optimização do valor de *mtry* usando MOLMAPs 15x15. Comparação entre diversos valores de *mtry* é feita através do valor de Concordância.

MOLMAPs 15×15 (Neurónio vencedor valor 10, neurónios vizinhos valor 3) obtidos através de mapeamento segundo 59 propriedades físico-químicas/topológicas			
<i>mtry</i>	Concordância conjunto de treino de 145 compostos	Concordância <i>Out of Bag</i> (conjunto de treino)	Concordância conjunto de validação de 49 compostos
10	0,993	0,966	0,959
25	0,993	0,966	0,959
50	0,993	0,959	0,959
75	0,993	0,966	0,959
100	0,993	0,959	0,959

Conjuntos de treino e validação.

Pode-se observar que a variação dos valores de concordância com o valor de *mtry* é praticamente nula, não se podendo afirmar inequivocamente qual o melhor valor de *mtry*. Como o modelo baseado em *mtry* = 25 é um dos que apresenta os melhores resultados este foi aplicado a um conjunto de teste de 45 reacções, correspondendo ao conjunto de teste de 45 compostos usados no estudo a). Os resultados estão na **Tabela III.23**.

Tabela III.23: Modelo *Random Forest* $mtry = 25$ usando MOLMAPs 15x15, estudo detalhado relativamente à capacidade de classificação, de acordo com o mecanismo reaccional (a-f), referente ao conjunto de treino, conjunto independente de validação e conjunto de teste.

		Conjunto de treino	<i>Out of Bag</i> (conjunto de treino)	Conjunto de validação de 49 reacções	Conjunto de teste de 45 reacções
Concordância		0,993	0,966	0,959	0,911
k		0,991	0,955	0,947	0,889
λ_B		0,989	0,944	0,933	0,879
Predictibilidade	a	1,000	1,000	1,000	1,000
	nº objectos	26	26	9	12
	b	1,000	0,944	1,000	1,000
	nº objectos	19	19	6	8
	c	1,000	0,982	1,000	1,000
	nº objectos	55	55	19	10
	d	1,000	0,947	0,750	1,000
	nº objectos	18	18	6	11
	e	1,000	1,000	1,000	1,000
	nº objectos	8	8	2	5
	f	0,947	0,889	1,000	-
	nº objectos	19	19	7	-
Sensibilidade	a	1,000	1,000	1,000	1,000
	nº objectos	26	26	9	12
	b	1,000	0,895	1,000	0,625
	nº objectos	19	19	6	8
	c	1,000	1,000	1,000	1,000
	nº objectos	55	55	19	10
	d	0,947	1,000	1,000	0,909
	nº objectos	18	18	6	11
	e	1,000	1,000	1,000	1,000
	nº objectos	8	8	2	5
	f	1,000	0,842	0,714	-
	nº objectos	19	19	7	-
nº objectos total		145	145	49	45

Definições de concordância, k, λ_B , predictibilidade e sensibilidade aparecem descritas na secção II-2.4 b) – Metodologia Computacional.

Analisando mais detalhadamente os resultados relativos ao conjunto de validação é possível observar que a grande maioria das classes apresentam valores elevados relativos à predictibilidade e sensibilidade, as excepções são a predictibilidade relativa à classe “d” (substituição nucleofílica aromática – 0,750) e a sensibilidade relativa à classe “f” (outras reacções – 0,714), dada a presença de grupos/funcionalidades não representadas no conjunto de treino.

Para o conjunto de teste observam-se elevadas predictibilidades mas as sensibilidades são um pouco mais baixas, principalmente para a classe “b” (reacção de grupo acilo – 0,625). Os resultados

de k , λ_B , predictibilidades e sensibilidades para o conjunto de teste foram obtidos a partir da tabela de contingência apresentada na **Tabela III.24**.

Tabela III.24: Tabela de contingência relativamente aos resultados obtidos com o conjunto de teste de 45 reacções.

		Classe experimental					Total linha
		a	b	c	d	e	
Previsão	a	12	0	0	0	0	12
	b	0	5	0	0	0	5
	c	0	0	9	0	0	9
	d	0	0	0	10	0	10
	e	0	0	0	0	5	5
	f	0	3	0	1	0	4
Total coluna		12	8	9	11	5	

Na diagonal principal aparecem representados aqueles objectos em que há correspondência entre as classes reaccionais (a-f) experimentais e previstas.

Tendo em conta a **Tabela III.24** pode-se observar que há quatro reacções mal classificadas, três reacções de classe “b” sendo atribuído pelo modelo classe “f” a todas elas, e um objecto de classe “d” que foi previsto como sendo classe “f”. As quatro reacções mal classificadas aparecem representadas na **Figura III.21**.

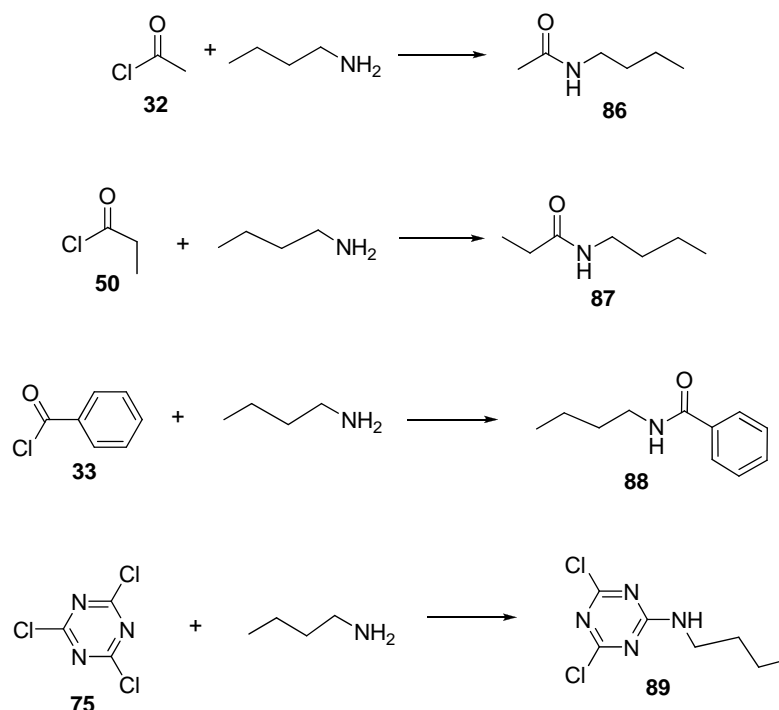
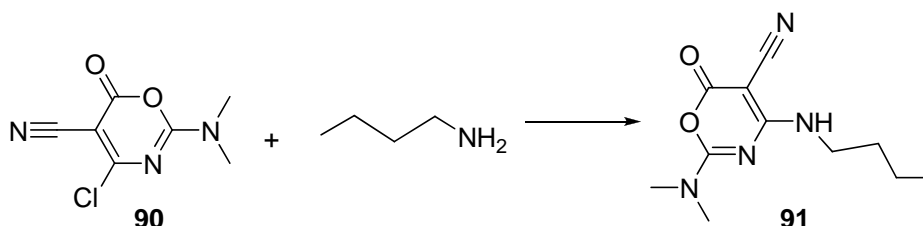


Figura III.21: Reacções mal classificadas segundo modelo *Random Forest*.

Pode-se observar que as três primeiras reacções envolvem cloretos de ácido que reagem segundo mecanismo “b” (Reacção de grupo Acilo). A quarta reacção representa o mecanismo “d” de substituição nucleofílica aromática. Relativamente às três reacções que ocorrem segundo mecanismo “b”, estas são classificadas pelo modelo como sendo classe “f” sendo esta a classe a apresentar maior probabilidade (0,584, 0,584 e 0,716 para as reacções das moléculas **32**, **50** e **33** respectivamente), a segunda classe mais provável é a classe correcta (classe “b”) com probabilidades compreendidas no intervalo 0,2 - 0,3. A razão para estes três objectos terem sido mal classificados está relacionada com o facto de que a reacção mais próxima no conjunto de treino ser de classe “f” (**Esquema III.9**):



Esquema III.9: Reacção contida no conjunto de treino mais próxima das reacções mal classificadas do conjunto de teste.

As reacções das moléculas **32**, **50** e **33** envolvem a quebra de ligação carbono-cloro, tal como o caso da molécula **90**, muito embora os ambientes químicos sejam diferentes. Ainda assim as ligações carbono-cloro dos três cloretos de ácido correspondem à mesma posição de MOLMAP ocupada por ligação carbono-cloro de molécula **90**. Desta forma não havia maneira do modelo distinguir os dois tipos de ambiente químico, sendo todos os objectos classificados segundo classe “f”.

O quarto objecto mal classificado é a reacção da molécula **75** (**Figura III.21**). Esta reage segundo mecanismo “d” (substituição nucleofílica aromática) mas o modelo classifica-a segundo classe “f”, sendo esta a classe claramente mais provável (1,000). Esta classificação incorrecta deve-se ao facto de a reacção mais próxima no conjunto de treino ser também a reacção da molécula **90** (**Esquema III.9**) que é de classe “f”. Em ambas as reacções está envolvida a quebra de uma ligação carbono-cloro, num dos casos o átomo de carbono é olefínico (molécula **90**) no outro caso é aromático (molécula **75**). Ambas as ligações carbono-cloro vão corresponder à mesma posição no MOLMAP. Da mesma forma estes dois objectos vão ter em comum a formação de ligação carbono-azoto simples correspondendo também à mesma posição em MOLMAP. Assim sendo não há distinção entre a reacção da molécula **90** e a reacção da molécula **75** por parte do modelo *Random Forest* sendo as duas reacções classificadas segundo mecanismo “f”, a proximidade entre estes dois objectos é 1,000. As quatro reacções incorrectamente classificadas pelo modelo exemplificam o facto de os MOLMAPs utilizados não terem capacidade para discriminar alguns ambientes químicos.

Apesar de tudo o modelo *Random Forest* obtido é bastante satisfatório no que à capacidade de previsão diz respeito.

De forma a poder interpretar quais as características estruturais mais relevantes na classificação de reacções químicas em que a butilamina intervém como nucleófilo, foi construída uma Árvore de Decisão construída com MOLMAPs 15x15 usando parâmetros de referência (**Figura III.22**).

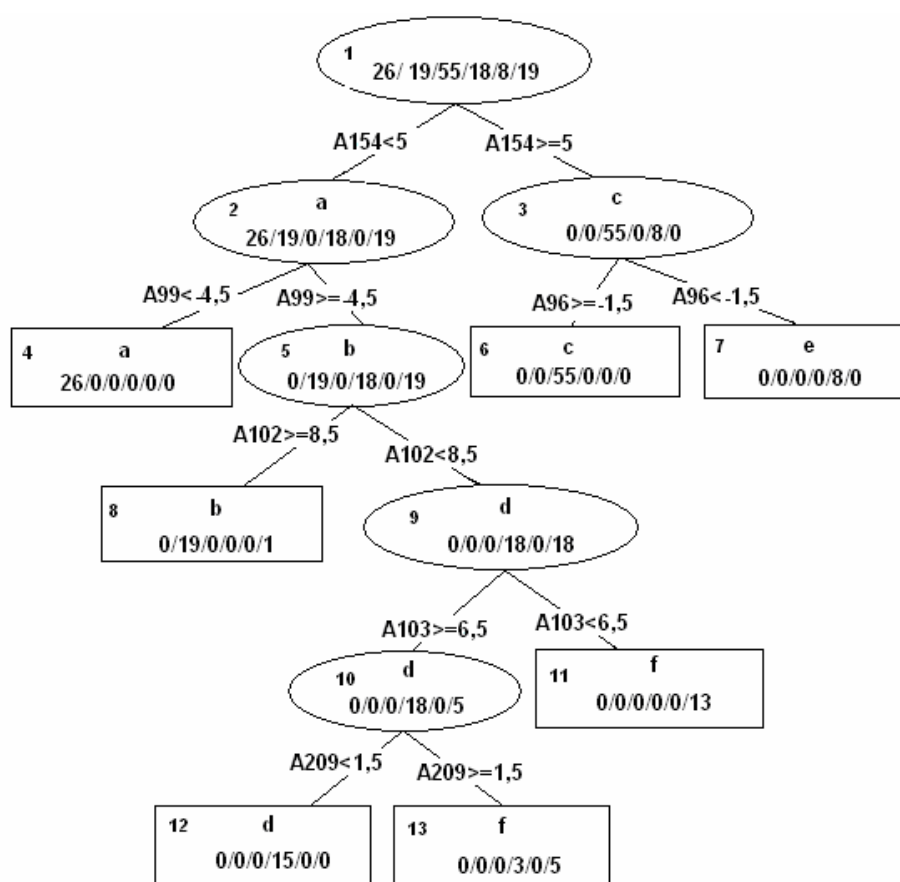


Figura III.22: Representação esquemática da Árvore de Decisão utilizada para classificar reacções químicas segundo mecanismo reaccional.

Através da análise detalhada à **Figura III.22** é possível observar uma primeira partição obtida através do descritor A154, sendo possível isolar no nó 3 duas classes, a classe “c”, com 55 reacções, e a classe “e” com 8 reacções. Nesse nó todos os objectos possuem, para o descritor A154, um valor igual ou superior a 5. O descritor A154 representa ligações azoto-carbono simples sendo este carbono alifático. Como para este nó todos os objectos aí contidos possuem descritor A154 superior a “0” na prática vai significar a formação de uma ligação carbono-azoto simples em cada um dos produtos de cada uma das reacções contidas no nó 3. O carbono dessa ligação é alifático. Os únicos tipos de

reação em que ocorre a formação de uma ligação com estas características são as reações de Substituição nucleofílica “c” e Adição de Michael “e”.

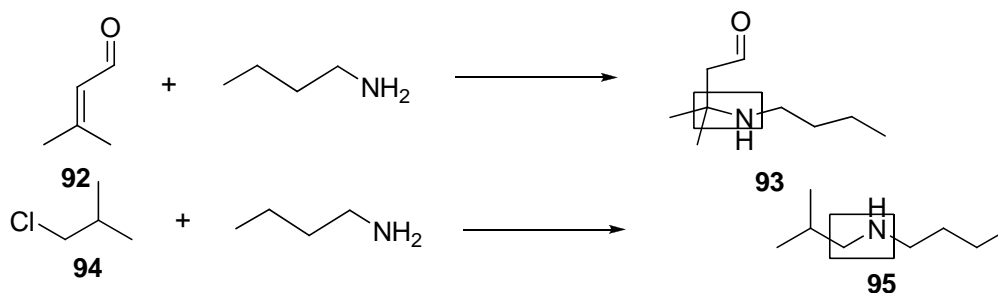
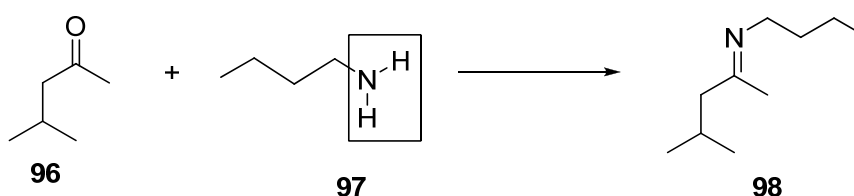


Figura III.23: Representação esquemática de reações de substituição nucleofílica e adição de Michael em que aparece evidente a formação de ligação azoto-carbono.

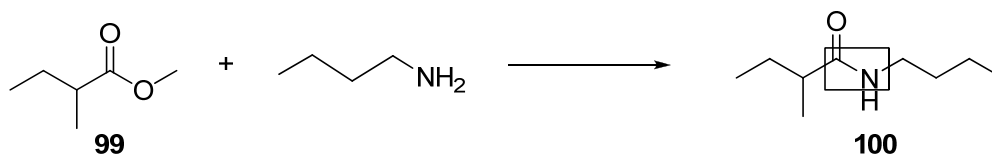
A partir do nó 3 dá-se nova partição através de descritor A96, este relacionado com ligação carbono-hidrogénio em olefina. Neste caso a partição para o nó descendente 7 (que também é nó terminal) é obtida para objectos com valor do descritor inferior a -1,5, o que representa o desaparecimento desse tipo de ligação durante a reação, logo adições de Michael “e” (**Figura III.23**). No outro nó descendente (nó terminal 6) são agrupadas as substituições nucleofílicas. Voltando a centrar atenções no nó 2, este vai sofrer partição a partir do descritor A99 que representa ligação azoto-hidrogénio. Desta partição resulta o nó descendente 4 (nó terminal) agrupando objectos em que o valor do descritor é inferior a -4,5. O valor negativo representa o desaparecimento de ligação azoto-hidrogénio (concretamente mais do que uma ligação para posição em MOLMAP correspondente a neurónio vizinho). Neste caso, o único tipo de reação aqui estudada que se baseia no desaparecimento de duas ligações azoto-hidrogénio são as reações do tipo “a” formação de base de Schiff (**Esquema III.10**) sendo isoladas no nó 4. No outro nó descendente, o nó 5, contendo objectos em que não há a perda de duas ligações azoto-hidrogénio são agrupadas as restantes três classes, “b”, “d” e “f”.



Esquema III.10: Representação esquemática de reação de formação de base de Schiff evidenciando a perda de duas ligações azoto-hidrogénio.

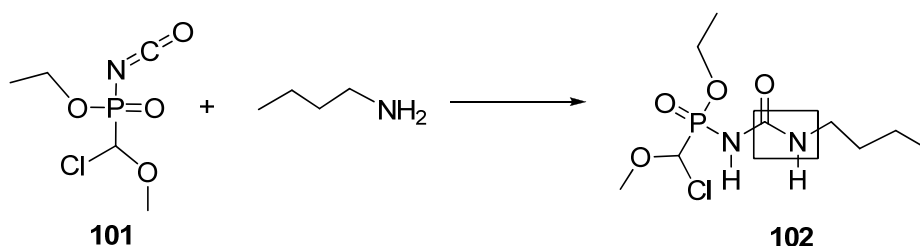
O nó 5, por sua vez, vai sofrer uma partição, operada a partir do descritor A102, que representa ligação carbono-azoto de amida, sendo obtido o nó 8, contendo objectos em que o descritor A102 é superior a 8,5 (aparecimento de ligação carbono-azoto de amida). Esse tipo de ligação é

formada em reacções do tipo “b”, reacção de grupo acilo (**Esquema III.11**), aparecendo praticamente isoladas no nó 8.



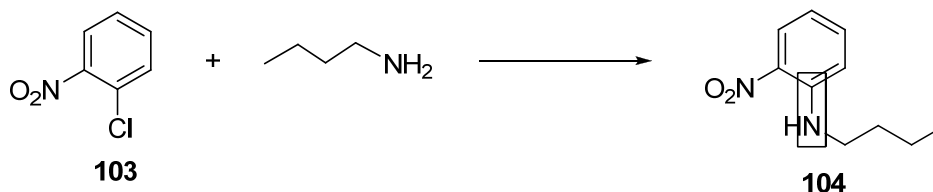
Esquema III.11: Representação de uma reacção de acilação resultando na formação de uma ligação carbono-azoto de amida, aparecendo aqui em evidência.

No nó 8 também aparece uma reacção de tipo “f” (outras reacções) que também consiste na formação de ligação carbono-azoto de ureia (**Esquema III.12**):



Esquema III.12: Reacção do tipo “f” classificada como sendo de tipo “b” onde também há a formação de uma ligação carbono-azoto de ureia.

No outro nó descendente, o nó 9, vão ser agrupadas os dois tipos de reacções ainda não isoladas, reacções de tipo “d” (Substituição Nucleofílica Aromática) e “f” (Outras Reacções) onde não há formação de ligação carbono-azoto de amida. O nó 9 vai sofrer partição a partir de descritor A103, que representa ligação carbono-azoto sendo o carbono aromático ou olefinico (este último em menos casos). Esta partição permite a obtenção de nó descendente 10, contendo objectos cujo descritor A103 é maior ou igual a 6,5, ou seja, este nó representa reacções em que há a formação de ligação carbono-azoto (carbono aromático ou olefinico) na conversão de reagente em produto (**Esquema III.13**):



Esquema III.13: Reacção de substituição nucleofílica aromática “d” evidenciando a formação de ligação carbono-azoto sendo o carbono aromático.

O nó **10** agrupa maioritariamente objectos de classe “d” (Substituição nucleofílica aromática) semelhantes à reacção representada no **Esquema III.13** e algumas reacções de classe “f”, onde também há a formação de ligação carbono-azoto mas com carbono olefínico –**Figura III.24**.

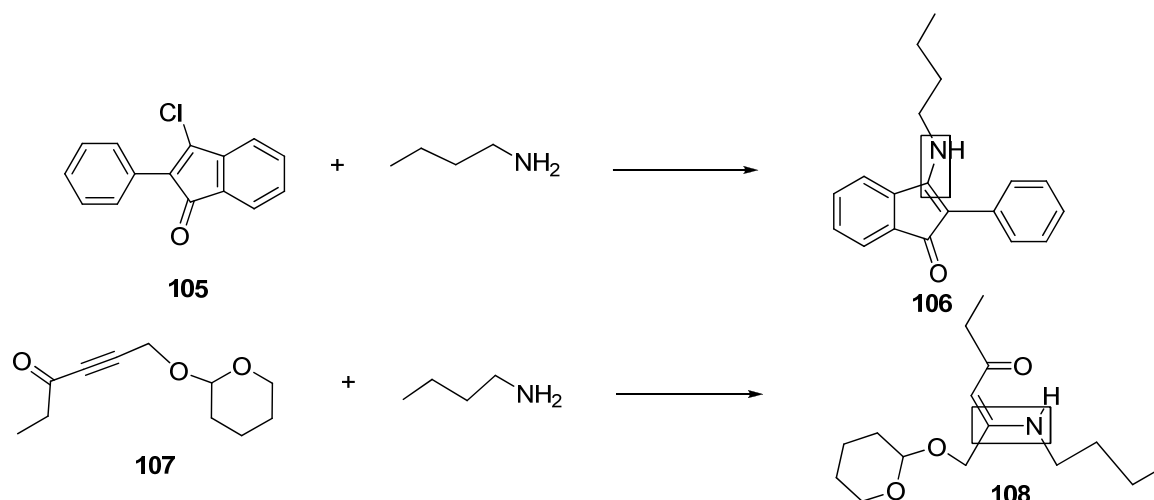


Figura III.24: Exemplos de reacções de classe “f”, que ainda ficaram associadas ao nó **10**.

O outro nó descendente (nó **11** terminal), relacionado com a ausência de formação de ligação carbono-azoto com carbono aromático ou olefínico, isola exclusivamente reacções de tipo “f”. Finalmente o nó **10**, que agrupa objectos de classe “d” e “f”, vai sofrer partição a partir do descritor A209, referente a ligação carbono-carbono olefínica possuindo grupo electroattractor como substituinte. Com esta partição é obtido o nó descendente terminal **12**, onde são agrupados objectos com o descritor inferior a 1,5 (isola objectos exclusivamente de classe “d”). No nó **13** estão agrupados maioritariamente objectos de classe “f” sendo exemplos as duas reacções representadas na **Figura III.24**. Ainda assim, aparecem neste nó terminal alguns objectos de classe “d”, que mesmo não contendo ligação olefínica possuindo grupo electro-atractor como substituinte, possuem ligações aromáticas possuindo grupo electro-atractor como substituinte permitindo somar determinado valor à posição A209 do MOLMAP respectivo. Fica assim evidente alguma incapacidade de discriminação de ligações por parte deste descritor. Ou seja, para a obtenção de uma árvore de decisão eficaz é necessário com poucos descritores separar todas as classes. Para isso acontecer é necessário que os descritores seleccionados tenham capacidade de discriminação mas ao mesmo tempo devem englobar uma série de ambientes químicos abrangente e nem sempre esse equilíbrio é perfeitamente obtido.

Para esse equilíbrio são necessários MOLMAPs de dimensões adequadas e um método de aprendizagem automático que opere adequadamente sobre esses MOLMAPs. O modelo Árvore de Decisão, não é o modelo ideal para operar sobre estes MOLMAPs de forma a obter classificações correctas. Ainda assim este modelo permite a obtenção de percentagem de acertos elevada para o conjunto de treino (concordância = 0,966) e para o conjunto de teste (concordância = 0,898). A

grande vantagem deste tipo de modelo é a sua interpretabilidade como ficou patente. O modelo RF, como foi visto anteriormente, é mais eficaz a classificar reacções, e ainda que não seja tão interpretável como o modelo Árvore de Decisão, é possível saber quais são os descritores mais importantes para a classificação de reacções. Estão listados na **Tabela III.25** os dez descritores mais importantes pelo modelo RF:

Tabela III.25: Os dez descritores segundo o modelo *Random Forest* – MOLMAP 15x15, *mtry* = 25 e respectivas medidas de importância.

Posição	Descritor	Importância (ver Capítulo I)
1	A154	1,892
2	A169	1,834
3	A103	1,563
4	A104	1,475
5	A119	1,455
6	A102	1,425
7	A73	1,360
8	A139	1,303
9	A152	1,284
10	A177	1,275

Os descritores A154, A102 e A103 tinham sido incluídos na árvore de decisão. Os descritores A169, A104, A119 e A139 apesar de não terem sido incluídos na árvore são vizinhos de descritores incluídos. O descritor, A169, representa ligação azoto-carbono, com carbono alifático e com azoto associado ou não a funcionalidade isotiocianato funcionalidade reactiva na presença de nucleófilo. Na quarta posição descritor A104, que representa ligação carbono-azoto, em que o carbono pertence a anel aromático e azoto pertence a grupo nitro. Já na quinta posição surge descritor A119, associado a ligações carbono-azoto simples e duplas conjugadas, a iminas e sendo iminas respectivamente. O oitavo lugar é ocupado por descritor A139, representado por ligações carbono-halogénio (carbono alifático). Os restantes descritores que não estão relacionados com os descritores seleccionados pela árvore de decisão são os descritores A73, A152 e A177. O descritor A73 surge na sétima posição, associado a ligação carbono-halogénio, com carbono a pertencer a anel aromático, estando também associado a ligação carbono-enxofre, o carbono pode ser olefínico e o enxofre pode pertencer a grupo electro-atractor sulfona. Já a nona posição é ocupada por descritor A152 que representa ligação carbono-azoto simples, estando o carbono ligado a olefina ou anel aromático. Finalmente, na décima posição surge descritor A177, que representa ligação carbono-azoto dupla, representando essencialmente iminas que principalmente surgem como produto de reacções do tipo “a”.

Nesta secção foi optimizado um modelo *Random Forest* e construído um modelo Árvore de Decisão, ambos os modelos apresentam excelente capacidade de classificação de reacções,

principalmente o modelo *Random Forest*. O modelo Árvore de Decisão é mais interpretável que modelo *Random Forest* sendo obtido no primeiro regras simples de atribuição de classe. Estes modelos podem ser utilizados no auxílio à construção de bibliotecas reaccionais mais extensas tendo em vista o estudo mais detalhado dos factores reaccionais/estruturais associados a reacções orgânicas em que intervém um nucleófilo como a butilamina.

III-3.3 c): Estudo de Reactividade de Compostos na Presença de butilamina

Nas duas secções anteriores o objectivo principal era a classificação de mecanismos reaccionais a partir de compostos (estudo **a**) e reacções (estudo **b**), usando um conjunto de reacções com butilamina, nucleófilo modelo. No estudo desta secção, à semelhança dos estudos **a**) e **b**), o ponto de partida foi a base de dados de reacções químicas em que intervém a butilamina, utilizaram-se também descritores moleculares MOLMAP, mas pretendeu-se aqui contruir modelos para prever se um composto reage ou não com butilamina.

Se nos estudos anteriores se estudaram exclusivamente objectos em que a reacção química se concretizava, neste estudo foram usados compostos que reagem e simulados compostos que não reagem. O objectivo deste estudo passa pela criação de modelos que discriminem objectos reactivos de objectos não reactivos, pondo em evidência os factores estruturais associados a essa discriminação. Tendo isto em consideração e sabendo que a reactividade química é regida principalmente por propriedades físico-químicas de ligações químicas, procurou-se usar exclusivamente neste estudo propriedades físico-químicas das ligações. Vai-se assim procurar distinguir, da forma mais genérica possível, ligações reactivas de ligações não reactivas.

Neste estudo, para o cálculo das propriedades físico-químicas das ligações foi usado o programa PETRA versão 4.0, surgindo na altura do início deste estudo como uma versão do PETRA melhorada. Adicionalmente esta versão do PETRA baseia o cálculo das cargas π em métodos teoricamente bem estabelecidos, como é o caso da teoria das orbitais moleculares de Hückel (HMO).

Partindo das 971 reacções seleccionadas foi calculada uma série de propriedades físico-químicas para cada ligação elegível (ver Metodologia Computacional). Desse conjunto total de ligações foram seleccionadas 4980 ligações químicas, que foram utilizadas para treinar grande multiplicidade de mapas auto-organizativos de dimensão 29×29. Cada mapa era treinado por ligações representadas por um conjunto específico de propriedades físico-químicas. No final foi seleccionado aquele conjunto de propriedades físico-químicas que permitia uma melhor discriminação de ligações iguais ou semelhantes contidas em ambientes químicos distintos. Este passo de selecção do conjunto de propriedades físico-químicas a representar cada ligação química é fundamental na obtenção de

MOLMAPs que codifiquem adequadamente a informação contida nesta base de dados reaccional de forma a obter bons modelos de previsão de reactividade química.

O passo de selecção de propriedades físico-químicas de ligação insere-se numa sequência de passos que no final conduz à obtenção do modelo de previsão. Essa sequência é a seguinte (ver mais detalhes na Metodologia Computacional): **i)** Estandardização de reacções, **ii)** selecção de reacções, **iii)** selecção de conjuntos de ligações para treinar redes de *Kohonen*, **iv)** selecção dos melhores conjuntos de propriedades físico-químicas de ligação química, **v)** normalização de descritores de ligações, **vi)** treino de mapas auto-organizativos (SOM) baseados em redes de *Kohonen* de diferentes dimensões, **vii)** obtenção de descritores MOLMAP, **viii)** obtenção de pseudocompostos que não reagem. **ix)** construção de conjuntos de dados derivados de bases de dados reaccionais e **x)** construção de modelos de classificação de reactividade química.

Para seleccionar o melhor conjunto de propriedades físico-químicas de ligações testaram-se diversos conjuntos de ligações químicas, de acordo com procedimento explicitado na Metodologia Computacional. No final o melhor conjunto está composto por 11 propriedades físico-químicas de ligação química que são as seguintes: **1)** carga π de átomo A de ligação A-B, **2)** carga π de átomo B, **3)** carga σ de átomo A, **4)** carga σ de átomo B, **5)** carga total de átomo A, **6)** carga total de átomo B, **7)** electronegatividade de orbital associada a ligação σ de átomo A, **8)** electronegatividade de orbital associada a ligação σ de átomo B, **9)** soma de todas as cargas σ trocadas em todas as interacções entre os átomos A e B da mesma ligação, **10)** estabilização por ressonância de carga positiva no átomo A e carga negativa no átomo B na quebra heterolítica de ligação A-B e finalmente, **11)** estabilização por ressonância de carga positiva no átomo B e carga negativa no átomo A na quebra heterolítica de ligação A-B.

A capacidade de discriminação proporcionada por este conjunto de 11 propriedades físico-químicas foi avaliada através do mapeamento de diversos conjuntos de ligações. O primeiro conjunto baseou-se em 15 carbonilos que foram mapeados em mapa auto-organizativo 29×29 (**Figura III.25**) e que corresponde aos carbonilos das estruturas apresentadas na **Figura III.26**.

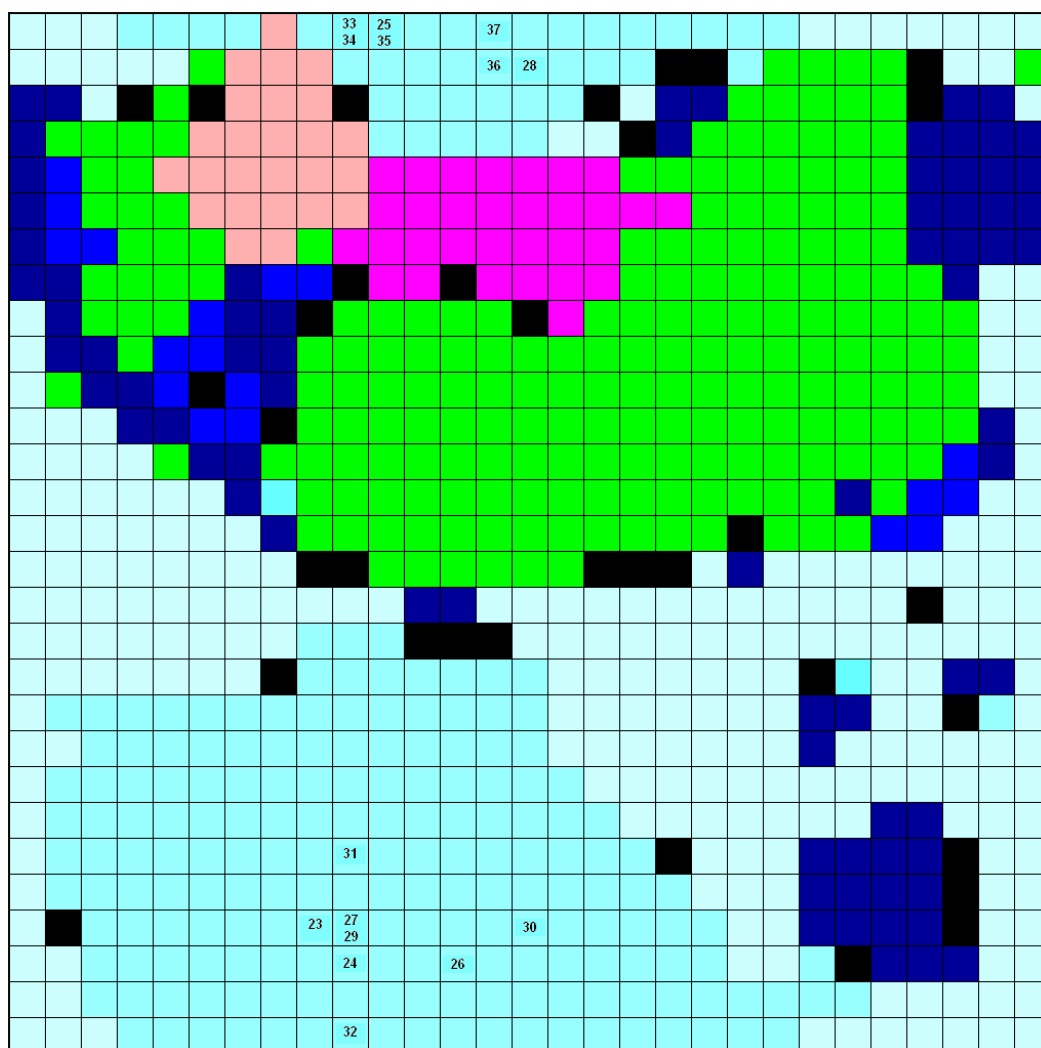


Figura III.25: Mapeamento de 15 carbonilos (C=O), referentes a estruturas representadas na Figura III.26 em SOM 29×29 baseado em 11 propriedades físico-químicas das ligações. Cor verde representa maioritariamente: ligações CC simples, duplas e triplas, CN simples, CSi simples, CH, ClC, BrC, CP simples, CS simples e ClS simples. Cor azul pálida representa ligações NC simples, dupla, tripla, NH, NO simples, NN simples e dupla e NP simples. Cor azul celeste: OC simples e dupla, OSi simples, OH simples, ON simples, OP simples e OS dupla. Cor azul: ligações ClC e BrC,. Cor azul escura: ligações SC simples, dupla, SN simples, SP dupla, SH simples. Cor rosa: ligações FC. Cor lilás: ligações HC.

De acordo com mapeamento apresentado na **Figura III.25**, são mapeados no mesmo neurónio carbonilos de cloreto de ácido aromático **33** e cloreto de ácido α,β insaturado **34**. Noutro neurónio estão mapeados carbonilos de ácido carboxílico alifático **25** e éster alifático **35**. Este mapa também não tem capacidade de discriminar carbonilos de cetona α,β insaturada **27** e cetona aromática **29** que aparecem mapeados no mesmo neurónio. Os restantes carbonilos estão mapeados em neurónios isolados.

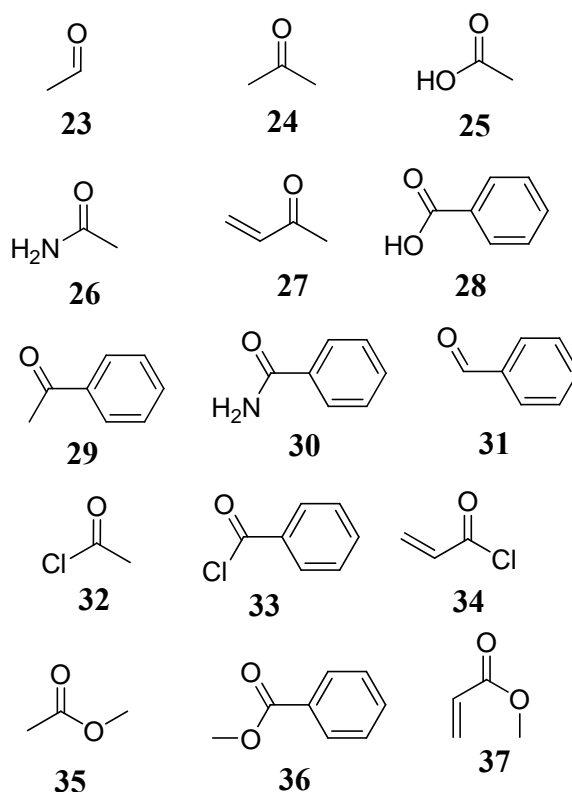


Figura III.26: 15 compostos contendo ligação carbonilo ($\text{C}=\text{O}$) constituindo diferentes funcionalidades e abrangendo diversos ambientes químicos.

À semelhança deste conjunto, outros conjuntos de ligações foram mapeados no mesmo mapa auto-organizativo 29×29 , aparecendo os resultados condensados na **Tabela III.26**. Percorrendo esta Tabela, pode-se observar que relativamente ao conjunto **b)** de ligações carbono-carbono, o mapa consegue distinguir ligações carbono-carbono simples, duplas, triplas e aromáticas, aparecendo todas mapeadas em neurónios distintos. Relativamente ao conjunto **c)** de ligações carbono-halogénio e carbono-hidroxilo o mapa distingue ligações carbono-flúor, carbono-cloro, carbono-bromo e carbono-hidroxilo aparecendo mapeadas em neurónios distintos. O mapa só não consegue discriminar entre ligações carbono-flúor aromáticas e olefinicas. Também não consegue distinguir ligações carbono-bromo alifáticas e aromáticas, assim como não consegue distinguir entre ligações carbono-cloro alifáticas e olefinicas. Relativamente ao conjunto **d)** de ligações o mapa consegue distinguir ligações carbono-azoto de amina, imina e nitrilo. Continuando a percorrer a **Tabela III.26**, observa-se que relativamente ao conjunto **e)** o mapa distingue ligações carbono-azoto de aminas e de grupos nitro. Já relativamente ao conjunto **f)** o mapa auto-organizativo consegue distinguir ligações carbono-oxigénio de grupos carbonilo e grupos hidroxilo. Por sua vez mapeado o conjunto **g)** observou-se distinção entre ligações carbono-azoto de aminas e amidas. Finalmente, mapeando o conjunto **h)** pôde-se verificar a distinção de diversas ligações carbono-carbono de esqueleto aromático estando associadas a diferentes grupos substituintes como é o caso do cloro, grupo nitro e hidrogénio.

Tabela III.26: Capacidade do mapa auto-organizativo para discriminar ligações representadas por um conjunto de 11 descritores físico-químicos.

Tipos de ligação	Em neurónios isolados/Separações	No mesmo neurónio
a) Carbonilos	<ul style="list-style-type: none"> - Ester α,β insaturado - Ester Aromático - Ácido Carboxílico Aromático - Aldeído Aromático - Aldeído Alifático - Cetona Alifática - Amida Alifática - Amida Aromática - Cloreto de Ácido Alifático 	<ul style="list-style-type: none"> - Cloreto de Ácido Aromático e Cloreto de Ácido α,β insaturado - Ácido Carboxílico Alifático e Éster Alifático - Cetona α,β insaturada e Cetona Aromática
b) ligações carbono-carbono, simples, duplas, triplas, Aromáticas	<ul style="list-style-type: none"> - Ligação CC simples -Ligação CC dupla -Ligação CC tripla -Ligação CC Aromática 	-
c) ligações carbono-X, em que X é halogénio ou hidroxilo	<ul style="list-style-type: none"> - Ligações CBr - Ligações CCl - Ligações CF - Ligações CHidroxilo 	<ul style="list-style-type: none"> - Ligações CF Aromáticas e Olefinicas - Ligações CBr Aromáticas e Alifáticas - Ligações CCl Alifáticas e Olefinicas
d) ligações carbono-azoto de amina, imina e nitrilo	<ul style="list-style-type: none"> - Amina - Imina -Nitrilo 	-
e) ligações carbono-azoto de grupos amina e nitro	<ul style="list-style-type: none"> - Amina - Nitro 	-
f) ligações carbono-oxigénio de hidroxilos e carbonilos	<ul style="list-style-type: none"> - Hidroxilos - Carbonilos 	-
g) ligações carbono-azoto em aminas e amidas	<ul style="list-style-type: none"> - Amina -Amida 	-
h) ligações carbono-carbono Aromáticas	<ul style="list-style-type: none"> - C=C - C=C-Cl - C=C-NO₂ 	-

Ligações discriminadas aparecem mapeadas em neurónios isolados, caso contrário aparecem mapeadas juntamente com outras ligações distintas no mesmo neurónio.

De acordo com resultados apresentados na **Tabela III.26** pode-se observar que este conjunto de 11 propriedades físico-químicas é bastante eficaz a separar as principais funcionalidades.

Com este conjunto de 11 propriedades físico-químicas foram treinados mapas auto-organizativos de outras dimensões para além de 29×29 (25×25, 20×20, 15×15, 10×10 e 7×7). Com estes mapas auto-organizativos foram obtidos MOLMAPs das respectivas dimensões para representar cada composto de cada conjunto de objectos de forma a treinar e testar um modelo de reactividade. Esse modelo visa prever se um determinado composto (Reagente) vai reagir na presença de butilamina. Para a construção desse modelo partiu-se de um conjunto de reacções no seguinte formato (**Esquema III.14**):



Esquema III.14: Representação genérica de uma reacção em que intervém a butilamina.

Este tipo de reacções inclui diversos mecanismos reaccionais associados a funcionalidades diferenciais. Os mecanismos reaccionais incluídos estão representados nos **Esquemas III.1-III.6** contidos na Introdução Teórica e referidos nos estudos **a)** e **b)** anteriores. Relativamente ao conjunto de treino, como era necessário para treinar tanto objectos que reagissem como objectos que não reagissem, e como na base de dados original só havia informação relativamente a compostos que reagiam, foi necessário gerar objectos que não iriam reagir. Estes objectos foram gerados a partir da base de dados reaccional gerando MOLMAPs para reagente e para o respectivo produto. Estes MOLMAPs foram comparados posição a posição de forma a identificar tipos de ligação que não sofriam alteração durante a reacção. A codificação desse tipo de ligações em MOLMAP resultante constituía um novo objecto, um pseudocomposto que não reagia. (Ver Metodologia Computacional).

Numa primeira abordagem, antes de comparar MOLMAP de reagente e respectivo produto, subtraiu-se ao MOLMAP do produto o MOLMAP da butilamina. Desta forma, no caso de uma amina presente no reagente sofrer alteração durante a reacção, esta não seria ocultada pela adição da butilamina ao gerar o produto. Construiu-se conjunto de treino com MOLMAPs de reagentes (que reagiam), presentes na base de dados reaccional, e com MOLMAPs de pseudocompostos (que não reagiam). Neste estudo variou-se a dimensão e tipo de MOLMAP.

Foram considerados dois tipos diferentes de MOLMAP. No primeiro caso, quando uma ligação de determinada molécula activasse neurónio vencedor, à respectiva posição do MOLMAP seria atribuído o valor 1. MOLMAPs com essas características estão identificados como sendo do tipo 1_0. No outro tipo de MOLMAP, quando determinada ligação activasse o neurónio vencedor, para

além de ser somado àquela posição do MOLMAP o valor 10, também às posições do MOLMAP correspondentes aos neurónios vizinhos seria somado determinado valor, neste caso 1. MOLMAPs com estas características estão identificados como sendo do tipo 10_1.

Nesta primeira abordagem, foram construídos modelos baseados em algoritmo *Random Forest*, mantendo constante o número de árvores = 1000 e *mtry* = 25, sendo obtidos os resultados expressos na **Tabela III.27**.

Tabela III.27: Comparação, relativamente ao valor de Concordância, entre os diversos modelos *Random Forest* construídos com MOLMAPs de diversos tipos e dimensões.

Dimensão do MOLMA P	MOLMAP tipo 10_1			MOLMAP tipo 1_0		
	Concordância conjunto de treino 1844 objectos	Conjunto de treino <i>Out of Bag</i>	Conjunto de validação 49 objectos	Conjunto de treino 1844 objectos	Conjunto de treino <i>Out of Bag</i>	Conjunto de validação 49 objectos
29×29	1,000	0,855	0,755	0,986	0,758	0,816
25×25	1,000	0,877	0,653	0,992	0,745	0,776
20×20	1,000	0,918	0,694	0,995	0,754	0,816
15×15	1,000	0,961	0,612	0,992	0,766	0,714
10×10	1,000	0,974	0,551	0,997	0,824	0,755
7×7	0,999	0,983	0,551	0,993	0,829	0,714

Conjuntos de treino e validação.

Tendo em conta os resultados apresentados na **Tabela III.27** para um problema binário, ou seja, em cada caso afirmar se vai ocorrer ou não reacção, é possível observar que todos os modelos construídos se ajustam bem às reactividades apresentadas por objectos do conjunto de treino (em todos os casos a concordância é superior a 0,980). Todos os modelos construídos foram avaliados na capacidade de previsão, sendo realizado ensaio *Out of Bag* com conjunto de treino. Nesta avaliação foi igualmente utilizado conjunto independente de validação de 49 compostos. Tendo em atenção esses dois itens pode-se observar que MOLMAPs de dimensões inferiores a 20×20 apresentam elevadas concordâncias para ensaios *Out of Bag* mas baixas concordâncias para o conjunto de validação. Este facto é particularmente evidente em MOLMAPs do tipo 10_1. O motivo para estes resultados está relacionado com o facto de MOLMAPs de pequenas dimensões terem maior proporção de posições de MOLMAP alteradas na conversão de reagentes em produtos, provocando aumento na proporção de posições de MOLMAP nulas para pseudocompostos gerados (que não reagem). Assim, os MOLMAPs referentes às duas classes consideradas (reage vs não reage) vão distinguir-se de uma forma artificialmente trivial. Desta forma vão ser obtidos elevados valores de concordância para ensaio *Out of Bag*, que não se reflectem nas previsões para moléculas reais como é o caso dos objectos do conjunto de validação. Os objectos do conjunto de validação são compostos presentes na base de dados reaccional inicial que se sabe que reagem, estando adicionalmente

incluídas moléculas baseadas em compostos presentes na base de dados reaccionais mas cujas funcionalidades reactivas foram previamente apagadas de forma a serem compostos que não reagem. Neste caso verifica-se ser indispensável o uso do conjunto de validação na optimização de modelo *Random Forest*.

De forma a compreender melhor o que se estava a passar com este tipo de modelos, já que as concordâncias obtidas para conjunto de validação não ultrapassavam 0,820, valor pouco satisfatório, foi construído um modelo Árvore de Decisão baseado em MOLMAPs 29×29 do tipo 1_0 (**Figura III.27**). Daqui em diante vão ser estudados MOLMAPs do tipo 1_0 e não 10_1 já que aqueles são mais facilmente interpretáveis e não apresentam uma diminuição tão marcada da concordância relativa ao conjunto de validação com a diminuição da dimensão do MOLMAP.

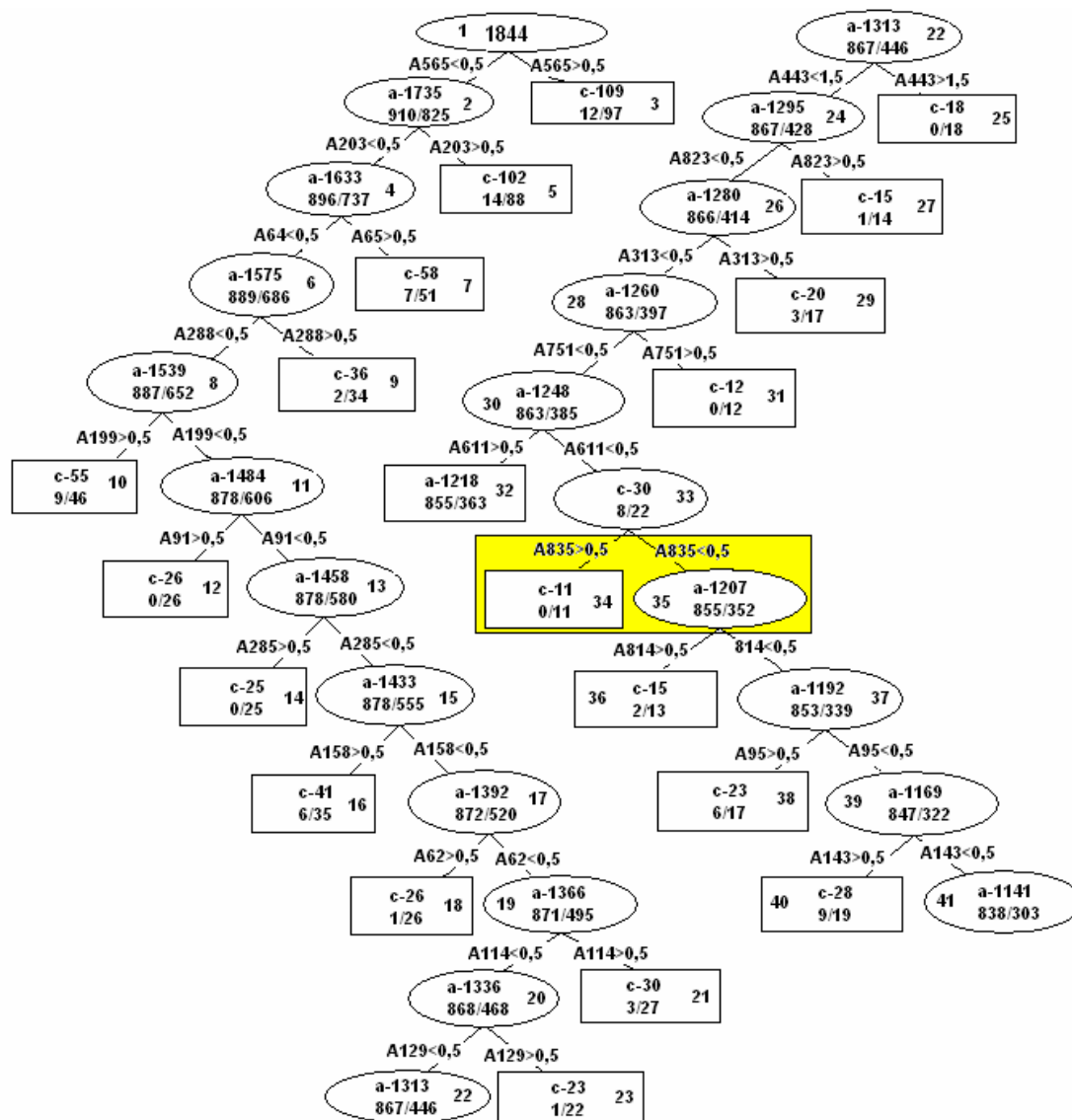


Figura III.27: Árvore de Decisão, relativa à classificação de compostos segundo a reactividade na presença de butilamina, construída com MOLMAPs 29×29 tipo 1_0 em que a geração de pseudocompostos que não reagem baseou-se na subtracção de MOLMAP da butilamina ao MOLMAP do Produto antes da comparação com MOLMAP de Reagente – classe “c” reage, classe “a” não reage. Assinalado a amarelo aparece descritor A835 associado a amins, funcionalidade que não está associada à reactividade na presença de nucleófilos.

Tendo em conta a Árvore de Decisão apresentada na **Figura III.27** é possível observar que o isolamento em nós terminais de objectos maioritariamente de classe “c” (reage) se rege pela presença de determinado tipo de ligação, representado pela condição “> - maior que” associada a determinado descritor MOLMAP. Esse tipo de condição faz todo o sentido visto que é a presença de determinados tipos de ligação que causa reactividade e não a ausência de determinado tipo de ligação que causa

inércia química. Na generalidade, os descritores MOLMAP seleccionados reportam-se a ligações tradicionalmente associadas a reacção com nucleófilo, como é o caso de ligações carbono-oxigénio simples de ésteres (A203), ligações carbono-halogénio associadas a anel aromático (A64), carbonilos cetónicos (A288) e de aldeídos (A285), tal como ligações carbono-halogénio alifáticas (A129) entre outras. Uma informação importante retirada desta árvore de decisão reside no facto de ter sido seleccionado o descritor A835 (assinalado a amarelo na **Figura III.31**). Este descritor representa ligações azoto-hidrogénio e azoto-carbono em aminas, ou seja, apesar de ser um tipo de ligação não associada à reactividade com nucleófilos foi seleccionada como tal. Esta situação deve-se ao facto de, na geração de pseudocompostos que não reagem, o MOLMAP de butilamina ser subtraído ao MOLMAP do produto. Em muitos casos as posições afectadas no MOLMAP resultante vão ter valores mais baixos do que os de posições equivalentes no MOLMAP do reagente respectivo. Ou seja, no MOLMAP gerado para o pseudocomposto, a essas posições vai ser atribuído o valor “0” já que a correspondente ligação é percebida como tendo desaparecido na conversão de reagentes em produtos. Neste contexto, quando o algoritmo da Árvore de Decisão está a avaliar quais as regras mais discriminatórias entre objectos da classe “reage - c” (onde há compostos que contêm funcionalidade amina e portanto valor não nulo para os correspondentes descritores, mas que reagem noutra funcionalidade) e objectos da classe “não reage - a” (que vão conter valor nulo em posição A835 referente a ligações de amina), vai perceber erradamente a funcionalidade amina como uma funcionalidade reactiva. Ou seja, incluindo butilamina como reagente a funcionalidade amina vai ser percebida como reactiva, o que não interessa ao modelo que pretende identificar reactividade no outro possível reagente, não na butilamina.

A Árvore de Decisão permitiu obter uma concordância de 0,650 para o conjunto de validação. Este problema também se reflecte em modelo *Random Forest* equivalente no qual o descritor A835 surge na sétima posição relativamente à importância dos descritores. Para resolver essa situação foram seguidas duas estratégias distintas. A primeira consistia em identificar as posições do MOLMAP correspondentes à butilamina (A271, A565, A834 e A835) e não contabilizar esses descritores na construção do modelo *Random Forest*. O modelo *Random Forest* assim obtido (1000 árvores e *mtry* = 25) permitiu um aumento dramático da concordância para o conjunto de validação (concordância = 0,920) assim como um valor em ensaio *Out of Bag* (concordância = 0,760). A outra estratégia seguida foi, na geração de pseudocompostos que não reagem, comparar directamente MOLMAP de reagente e MOLMAP do produto respectivo sem intervir o MOLMAP da butilamina. Desta forma foram gerados MOLMAPs de diversas dimensões do tipo 1_0 para construção e avaliação da capacidade de previsão de modelos *Random Forest* (*mtry* = 25 e 1000 árvores) respectivos.

Tabela III.28: Comparação, relativamente ao valor de concordância, entre os diversos modelos *Random Forest* construídos com MOLMAPs de diversas dimensões.

Dimensão do MOLMAP	MOLMAP tipo 1_0		
	Concordância conjunto de treino 1844 objectos	Concordância conjunto de treino <i>Out of Bag</i>	Concordância conjunto de validação 49 objectos
29×29	0,984	0,743	0,898
25×25	0,991	0,745	0,857
20×20	0,978	0,741	0,898
15×15	0,992	0,750	0,837
10×10	0,995	0,791	0,673
7×7	0,975	0,738	0,857

Conjuntos de treino e validação.

De acordo com resultados apresentados na **Tabela III.28**, esta estratégia também permitiu uma melhora muito significativa da concordância obtida com conjunto de validação, especialmente com MOLMAPs de dimensões 29×29 e 20×20 (concordância = 0,898 para ambos). Considerando ambas as dimensões de MOLMAP foi optimizado o valor de *mtry* (**Tabela III.29**).

Tabela III.29: Optimização do valor de *mtry* usando MOLMAPs 29×29 e 20×20. Comparação entre diversos valores de *mtry* é feita através do valor de concordância.

<i>mtry</i>	MOLMAP dimensão 29×29 tipo 1_0			MOLMAP dimensão 20×20 tipo 1_0		
	Concordância conjunto de treino 1844 objectos	Concordância conjunto de treino <i>Out of Bag</i>	Concordância conjunto de validação 49 objectos	Concordância conjunto de treino 1844 objectos	Concordância conjunto de treino <i>Out of Bag</i>	Concordância a conjunto de validação 49 objectos
25	0,984	0,743	0,898	0,978	0,741	0,898
50	0,995	0,755	0,918	0,999	0,760	0,939
75	0,997	0,759	0,918	1,000	0,763	0,918
100	0,998	0,762	0,898	1,000	0,762	0,898
200	1,000	0,770	0,857	1,000	0,763	0,857
300	1,000	0,776	0,878	1,000	0,759	0,878

Conjuntos de treino e validação.

Tendo em conta resultados apresentados na **Tabela III.29** foi encontrado um modelo optimizado, aquele que apresentava valor mais elevado de concordância para conjunto de validação. Neste caso o modelo escolhido baseia-se em MOLMAPs 20×20 (tipo 1_0) usando *mtry* = 50 como parâmetro da *Random Forest*. Este modelo foi posteriormente testado, sendo obtidas previsões para conjunto independente de teste de 49 compostos não usados até aqui. Os resultados detalhados

referentes a este conjunto e aos conjuntos de treino e validação aparecem condensados na **Tabela III.30**.

Tabela III.30: Valores de Concordância, especificidade e sensibilidade obtidos por melhor modelo RF para conjuntos de treino, validação e teste.

Modelo <i>Random Forest</i> 1000 árvores, <i>mtry</i> =50 treinado com MOLMAPs dimensão 20×20 tipo 1_0			
	Conjunto de treino 1844 objectos	Conjunto de validação 49 objectos	Conjunto de teste 49 objectos
Concordância	0,999	0,939	0,878
Especificidade	0,999	0,958	0,960
Sensibilidade	0,999	0,920	0,792

É possível observar para o conjunto de teste um valor de concordância consideravelmente elevado (concordância = 0,878) correspondendo a seis objectos mal classificados, incluindo cinco falsos negativos e um falso positivo, daí resulta a discrepância relativamente ao valor da sensibilidade (0,792) e especificidade (0,960). A razão para as previsões incorrectas está relacionada com a falta de exemplos similares no conjunto de treino (evidenciado pela baixa proximidade verificada a objectos do conjunto de treino). Quatro das previsões problemáticas estão associadas a probabilidades perto de 0,500. Os outros dois compostos são falsos negativos. Um dos compostos possui uma funcionalidade não existente no conjunto de treino, o outro composto possui uma subestrutura reactiva que não reage no objecto mais próximo do conjunto de treino. Este objecto possui outra funcionalidade que reage preferencialmente. A associação entre as probabilidades de reacção designadas e as classes de reactividade experimental para conjunto de teste aparecem ilustradas na **Figura III.28**.

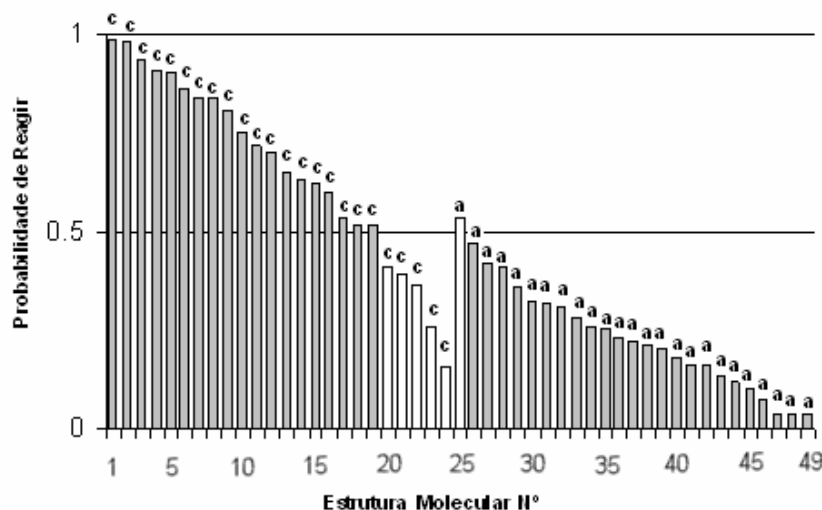


Figura III.28: Associação entre a reactividade experimental (c – reactivo, a – não reactivo) e a probabilidade de reacção designada pela RF para os 49 compostos do conjunto independente de teste. Previsões correctas assinaladas por barras a cinzento e previsões incorrectas assinaladas por barras brancas.

A partir dos valores de sensibilidade e especificidade obtidos para diversos patamares de probabilidade, considerando o conjunto independente de teste, foi gerada uma curva ROC. Esta é uma maneira de avaliar se o incremento na fracção de positivos verdadeiros não está relacionada com o incremento do número total de previsões positivas independentemente de serem verdadeiras ou falsas (Figura III.29). Esta medida de não correlação é a área entre a curva ROC e a linha de não discriminação. Essa área aparece representada pelo valor 0,930.

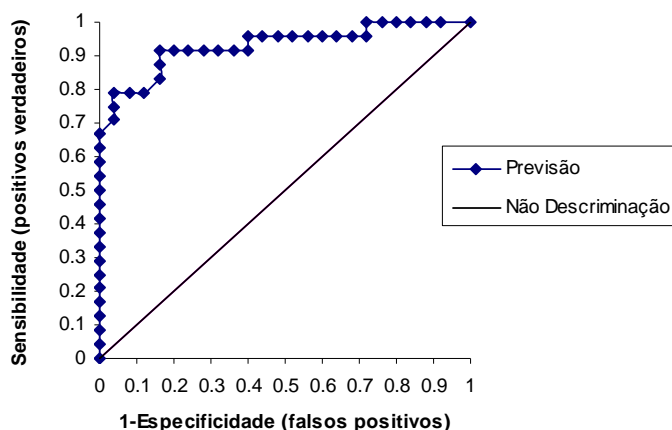
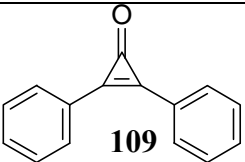
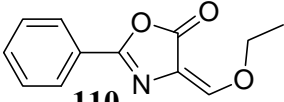

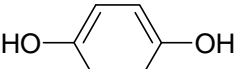
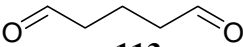
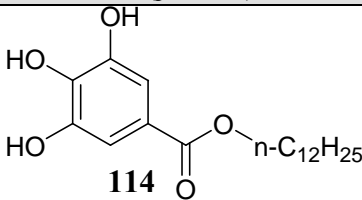


Figura III.29: Curva ROC obtida com os resultados do melhor modelo RF para os 49 compostos do conjunto independente de teste.

Este modelo final foi adicionalmente aplicado a novo conjunto de 32 compostos, e as previsões foram comparadas com resultados de reactividade *in vitro* na presença de péptido contendo lisina. A cadeia lateral da lisina, responsável pela nucleofilicidade desse péptido, é semelhante à butilamina e a reactividade na presença deste péptido correlaciona bem com o potencial de um composto para causar dermatite de contacto alérgica.³ A reactividade neste ensaio *in vitro* foi medida como a percentagem de péptido que reagiu com determinado composto. O modelo RF previu que 58.3% dos compostos com percentagem de reacção > 10% reagem, e que 60% dos compostos com percentagem de reacção < 10% não reagem (**Tabela III.31**). A concordância para este conjunto é 0,594.

Tabela III.31-1: Lista de compostos testados *in vitro* na presença de péptido contendo lisina

		
0,7 %	49,6%	91%
A (experimental)	c (experimental)	c (experimental)
a (previsão)	c (previsão)	a (previsão)
		
51,1%	85,4%	8,7%
c (experimental)	c (experimental)	c (experimental)
a (previsão)	c (previsão)	c (previsão)

Está incluída a percentagem de péptido contendo lisina que reagiu, a classe de reactividade experimental e a classe de reactividade prevista (c – reactivo, a – não reactivo).

Tabela III.31-2: Continuação da lista de compostos testados *in vitro* na presença de péptido contendo lisina.

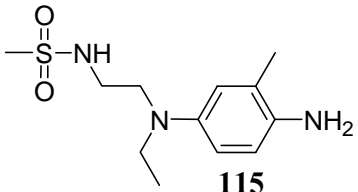
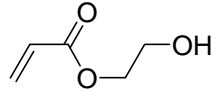
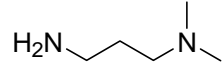
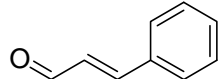
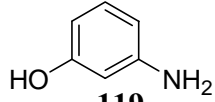
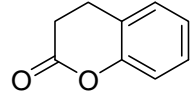
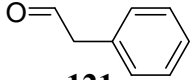
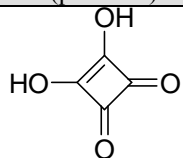
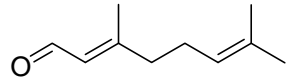
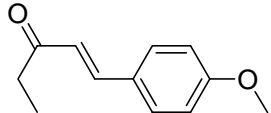
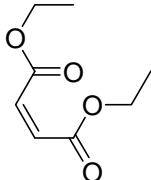
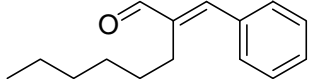
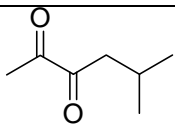
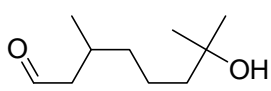
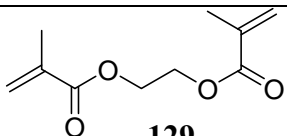
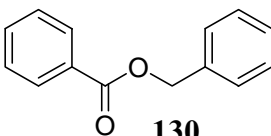
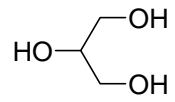
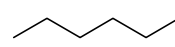
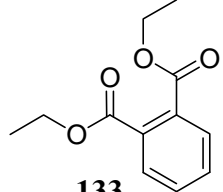
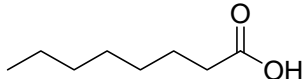
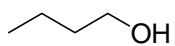
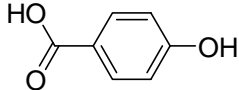
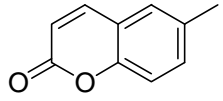
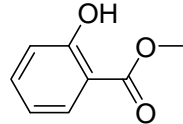
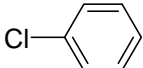
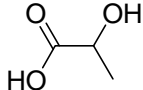
 115	 116	 117
13,6%	88,9%	-1,3%
c (experimental)	c (experimental)	a (experimental)
c (previsão)	a (previsão)	a (previsão)
 118	 119	 120
43,2%	1,2%	7,5%
c (experimental)	a (experimental)	a (experimental)
c (previsão)	a (previsão)	c (previsão)
 121	 122	 123
22,6%	4,8%	16,9%
c (experimental)	c (experimental)	c (experimental)
c (previsão)	c (previsão)	c (previsão)
 124	 125	 126
14,3%	85,5%	-1,6%
c (experimental)	c (experimental)	a (experimental)
c (previsão)	a (previsão)	c (previsão)

Tabela III.31-3: Continuação da lista de compostos testados *in vitro* na presença de péptido contendo lisina.

 127	 128	 129
7,5%	6,5%	12,4%
a (experimental)	a (experimental)	c (experimental)
c (previsão)	c (previsão)	a (previsão)
 130	 131	 132
3%	2,1%	-5,1%
a (experimental)	a (experimental)	a (experimental)
a (previsão)	a (previsão)	a (previsão)
 133	 134	 135
-0,7%	0,9%	1,2%
a (experimental)	a (experimental)	a (experimental)
a (previsão)	c (previsão)	a (previsão)
 136	 137	 138
2,2%	4%	1,6%
a (experimental)	a (experimental)	a (experimental)
c (previsão)	c (previsão)	a (previsão)
 139	 140	
1,3%	0,8%	
a (experimental)	a (experimental)	
a (previsão)	c (previsão)	

Tendo em conta os resultados apresentados na **Tabela III.31** é possível observar que três dos quatro falsos positivos são ácidos carboxílicos – **134**, **136** e **140**. A estes três compostos é associada a previsão “c – reage” já que no conjunto de treino existem ácidos carboxílicos que reagem com a butilamina como é o caso do ácido octadecanóico, ácido benzóico e ácido 2,2-dimetilpropiónico. Esta

discrepância de reactividades deve-se certamente às diferentes condições reaccionais existentes no teste *in vitro* e nas reacções em que se baseia o conjunto de treino. Por outro lado os dados que se utilizaram neste estudo não continham informação relativamente ao rendimento nem velocidade das reacções, podendo estar incluídos reagentes que conduzem a rendimentos muito baixos ou a reacções muito lentas.

O outro falso positivo é o composto **126**, um aldeído extremamente conjugado contendo uma longa cadeia alquílica em posição α que reduz a electrofilicidade do composto e aumenta o impedimento estéreo. Este composto reage de acordo com mecanismo de adição de Michael.⁴⁴ As moléculas do conjunto de treino mais próximas a **126** são o benzaldeído e **141**.

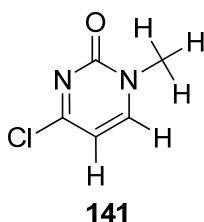


Figura III.30: Composto 141 que reage segundo mecanismo de substituição conjugada, semelhante ao mecanismo de adição conjugada representada pela adição de Michael.

Ésteres conjugados também são problemáticos, resultando em três falsos negativos – dois deles são as moléculas **125** e **129**. No conjunto de treino, os objectos mais próximos são dois pseudocompostos derivados de ésteres conjugados que incluem mais grupos funcionais reactivos, de acordo com a base de dados original, sendo descritos como reactivos mas permanecendo a funcionalidade éster intacta por selectividade. Este problema de selectividade aparece associado ao método de geração de pseudocompostos que não reagem, sendo gerados a partir de sub-estruturas que não reagem em compostos reactivos.

O composto **116** é um aceitador de Michael no processo de sensibilização cutânea²⁶ sendo igualmente reactivo no teste *in vitro* na presença de péptido contendo lisina³. O modelo RF considera o composto não reactivo já que o composto mais próximo no conjunto de treino é um pseudocomposto que não reage baseado num reagente contendo em simultâneo as funcionalidade isotiocianato e éster α,β insaturado. Este composto reage selectivamente no grupo isotiocianato, deixando a funcionalidade éster α,β insaturado codificada no MOLMAP do pseudocomposto. Novamente a questão da selectividade associada ao método de extração de informação negativa aparece evidenciada. A previsão obtida para este composto é no entanto obtida com uma probabilidade baixa (*ca* de 0,6). Outra explicação para as pobres previsões associadas aos ésteres deve-se à exclusão de reacções que não estão estequiometricamente equilibradas (tolerância máxima de perda de 6 átomos no produto em

relação aos reagentes). Diversas aminólises de ésteres, cujo álcool abandonante é maior que o metanol, foram excluídas (estes grupos abandonantes geralmente não aparecem representados na base de dados original como produtos).

Finalmente **111** e **112** são incorrectamente previstos pelo modelo como não reactivos. Ao composto **112** é associada a previsão “não reactivo - a” já que fenóis geralmente não são compostos reactivos na base de dados. De facto este é o único falso negativo com uma probabilidade longe de 0,5. O composto **112** reage provavelmente devido a oxidação com ar, prévia à reacção com a lisina. O modelo RF não contabiliza qualquer mecanismo de activação (seja ela metabólica ou não metabólica). Por sua vez o composto **111** é um aceitador de Michael extremamente potente na sensibilização cutânea, possuindo uma elevada conversão no ensaio *in vitro* na presença de péptido contendo lisina.³ O modelo RF considera o composto não reactivo já que um dos objectos do conjunto de treino mais semelhantes é um pseudocomposto que não reage derivado de **142**. Este composto reage através de mecanismo de substituição conjugada, semelhante à adição conjugada, representada por adição de Michael. Assim, o modelo aprende que este tipo de subestrutura não reage. Este resultado deve-se em grande medida à escassa presença de aceitadores de Michael na base de dados original usada na construção do modelo. Este resultado deve-se também às limitações no método de derivação de informação negativa já que algumas funcionalidades reactivas também estão codificadas em MOLMAP de pseudocompostos que não reagem.

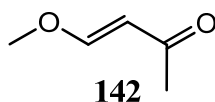


Figura III.31: Estrutura usada na derivação de MOLMAP de pseudocomposto que não reage semelhante a MOLMAP de 112.

Por sua vez, o composto **120** reage através de funcionalidade éster no processo de sensibilização cutânea²⁶ mas de acordo com resultado publicado por Gerberick *et al* em teste *in vitro*,³ este composto está situado abaixo do patamar para ser considerado composto reactivo. A previsão deste modelo relativamente a este composto é que reage estando a probabilidade de previsão perto de 0,7 e o composto mais próximo no conjunto de treino reage através do mesmo mecanismo sendo um éster activado também.

O composto **127** por sua vez reage através de formação de base de Schiff no processo de sensibilização cutânea sendo um sensibilizador fraco principalmente devido a impedimento estérico à volta de ambos os carbonilos¹ sendo um composto não reactivo em teste *in vitro* na presença de péptido contendo lisina.³ Segundo a previsão obtida por modelo RF o composto é reactivo, com o

composto mais próximo a ser a 1,2-ciclohexanodiona, que forma também base de Schiff na presença de butilamina, correspondendo a um elevado valor de proximidade em relação a **127**. Por sua vez, a probabilidade da previsão está perto de 0,6.

Relativamente ao composto **128** este é um sensibilizador cutâneo envolvido na formação de base de Schiff²⁶ mas que não reage *in vitro* na presença de péptido contendo lisina³ de acordo com o patamar de reactividade estabelecido nesse trabalho. Este composto é reactivo de acordo com o modelo RF, sendo o octanal o composto mais próximo do conjunto de treino, o qual também conduz à formação de base de Schiff na presença de butilamina, correspondendo a um valor elevado de proximidade. A probabilidade de previsão está acima de 0,9.

Finalmente, o composto **137** não é sensibilizador cutâneo de acordo com Ashby *et al.*²⁷ Este facto é reforçado por resultado obtido por Gerberick em ensaio *in vitro*³ estando abaixo do patamar estabelecido para ser considerado composto reactivo. De acordo com modelo de previsão, este composto reage, ainda que a probabilidade de previsão seja baixa (*ca* de 0,6). Curiosamente, o composto mais próximo no conjunto de treino é um pseudocomposto derivado de **143**, correspondendo a valor reduzido relativamente à proximidade. Na base de dados reaccional **143** conduz à formação de base de Schiff já que a funcionalidade enol pode ser tautomerizada a grupo carbonilo. De forma a obter pseudocomposto, a esta funcionalidade é atribuído valor “0” no MOLMAP correspondente, permanecendo outras funcionalidades (éster) que são consideradas “não reactivas”.

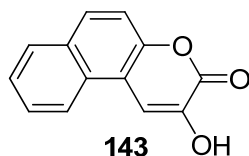


Figura III.32: Estrutura usada na obtenção de MOLMAP de pseudocomposto que não reage, semelhante a MOLMAP de **137**.

Relativamente a este conjunto, estudou-se também a influência do patamar mínimo de probabilidade de previsão para contabilizar objectos no cálculo do valor da concordância. O resultado desse estudo aparece representado na **Figura III.33**.

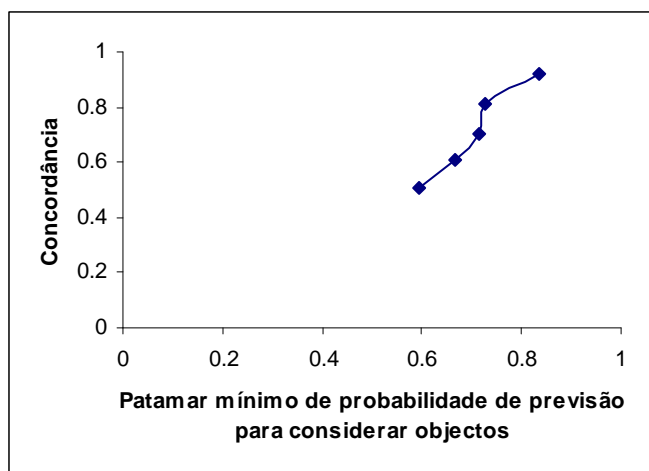


Figura III.33: Concordância vs patamar mínimo de probabilidade de previsão para considerar objectos de conjunto de 32 compostos testados *in vitro* por Gerberick *et al.*³

Adicionalmente este modelo foi validado atribuindo classe experimental a cada objecto do conjunto de treino de forma aleatória mantendo a proporção entre objectos classificado como reactivos e não reactivos em 1:1. A *Random Forest* treinada com este conjunto de treino modificado proporcionou valores de concordância baixos (0,490 e 0,612 para conjunto de validação e conjunto de teste respectivamente).

Relativamente à avaliação de quais os descritores mais importantes e qual a forma de serem interpretados, antes de analisar em pormenor o modelo RF, foi construída uma Árvore de Decisão. Este algoritmo tem algumas semelhanças com a *Random Forest* e é muito mais facilmente interpretável que esta última. O esquema da Árvore de Decisão construída aparece representada na **Figura III.34**.

Tabela III.32: Tipos de ligação representadas em árvore de decisão da Figura III.34.

Descritor	Tipo de ligação representada
A67	Ligação C-Cl em ambiente alifático, associada a anel aromático e cloreto de ácido
A274	Ligação C-O associada a fracção álcool de ésteres
A128	Ligação C-C em que um dos carbonos pertence a carbonilo de cetona
A244	C-S em tiol associado a anéis aromáticos
A110	C-C em que um dos carbonos pertence a carbonilo de ácido carboxílico ou éster / C-P associado a funcionalidade fosfonato
A129	C-C em que carbono pertence a carbonilo de cetona / C-Cl em que carbono é aromático
A45	C-C ou C-N em que carbono é aromático ou olefinico e o outro elemento é grupo electroattractor como nitro ou nitrilo
A303	C-O de epóxidos
A134	-N=C=P- e N=C=S
A354	Carbonilo de cetonas e amidas
A371	Carbonilo de ésteres e ácidos carboxílicos
A47	C-Cl de cloretos de ácido
A99	C=N em anel heteroaromático
A210	C-S, carbono aromático e enxofre pertence a tiol

Tendo em conta os descritores representados na **Tabela III.32** é possível observar que a grande maioria dos tipos de ligação seleccionados são efectivamente ligações reactivas na presença de nucleófilo como a butilamina, desde ligações C-Cl (A67), C-O de ésteres e epóxidos (A274 e A303 respectivamente), ligações associadas a carbonilos de cetona, éster e ácido carboxílico (A128, A129, A110, A354 e A371), ligações C-S em que o enxofre pertence a tiol, grupo de saída quando associado a anéis aromáticos em base de dados reaccional original (A244 e A210), ligações C-Cl de cloretos de ácido (A47) e ligações associadas a funcionalidades electrófilas (A134). O único descritor que à partida podia não fazer muito sentido representa ligações C=N em anel heteroaromático (A99), pelo menos se for interpretado isoladamente. Este descritor é precedido pelo descritor A371 na árvore de decisão da **Figura III.34** associado a ésteres e ácidos carboxílicos. Ou seja, ésteres e ácidos carboxílicos aromáticos não são reactivos de acordo com a base de dados reaccional original.

A discussão sobre esta Árvore de Decisão pode ajudar na interpretação dos descritores definidos como os mais importantes segundo o melhor modelo *Random Forest*, até porque alguns

descritores são os mesmos que foram seleccionados pela Árvore de Decisão. Os dez descritores mais importantes segundo o melhor modelo RF aparecem representados na **Tabela III.33**.

Tabela III.33: Os dez descritores mais importantes segundo melhor modelo RF – 1000 árvores, $mtry = 50$ obtido com MOLMAPs 20x20 do tipo 1_0 e respectivos valores de importância.

Posição	Descritor	Importância (ver Capítulo I)
1	A244	1,100
2	A67	1,080
3	A303	0,965
4	A214	0,935
5	A371	0,921
6	A274	0,907
7	A128	0,904
8	A312	0,902
9	A45	0,892
10	A313	0,885

Segundo resultados apresentados em **Tabela III.33**, sete dos dez descritores mais importantes também foram seleccionados pelo algoritmo Árvore de Decisão. As únicas excepções são os descritores A312, A313 e A214. O descritor A214 colocado na quarta posição representa ligações carbono-cloro em que o carbono pertence a anel aromático. Este tipo de ligação é quebrada em reacções de substituição nucleofílica aromática na presença de nucleófilo como é o caso da butilamina. Por sua vez o descritor A312, colocado na oitava posição, representa carbonilo de cetona, funcionalidade reactiva na presença de nucleófilo formando base de Schiff na presença de nucleófilo butilamina. Finalmente o descritor A313 colocado na décima posição representa carbonilos de cetonas α,β insaturadas, cetonas e aldeídos, todas estas funcionalidades são reactivas na presença de nucleófilo.

Tanto os modelos RF como Árvore de Decisão baseavam-se em conjunto alargados de descritores, neste caso MOLMAPs completos (por exemplo a MOLMAPs 29x29 correspondem 841 descritores). Estes algoritmos lidam bem com um conjunto grande de descritores inicial visto operarem de forma sequencial e seleccionarem apenas alguns descritores para os modelos construídos. Com outros algoritmos, como redes neuronais de *Kohonen* e *Support Vector Machines*, é conveniente uma selecção prévia de descritores.

Tendo em conta as especificidades associadas ao estudo contido nesta secção, foi testada uma nova forma de seleccionar descritores. Partindo de mapa auto-organizativo de dimensões 29x29 (representado na **Figura III.25**) foram aí mapeadas todas as ligações dos 922 reagentes e dos

respectivos 922 produtos presentes na base de dados seleccionada para este estudo. Cada ligação possuía uma etiqueta relativamente à sua pertença a um reagente (valor -1) ou um produto (valor 1). Nesse mapeamento foram seleccionados aqueles neurónios (posições do mapa) que possuíam maior proporção de ligações de reagentes. Pretendia-se assim seleccionar descritores associados a tipos de ligação mais tipicamente reactivas na presença de butilamina. Das 841 posições do mapa (29×29) foi seleccionado um subconjunto, correspondendo cada uma dessas posições a um determinado descritor MOLMAP de dimensão 29×29. O critério para selecção de neurónios foi a activação por pelo menos quatro ligações sendo maioritariamente ligações de reagente – 64 descritores. Partindo de MOLMAPs 29×29 de tipo 1_0 e 10_1, foram testados dois subconjuntos de descritores de cada tipo para treinar redes de *Kohonen* (totalidade dos 841 descritores MOLMAP e selecção de 64 descritores). Foram usadas inicialmente redes de *Kohonen* de dimensão 20×20 (**Tabela III.34**).

Tabela III.34: Comparação, relativamente a valores de Concordância, entre modelos baseados em redes de *Kohonen* 20x20 – baseados em MOLMAPs 29x29 do tipo 1_0 e 10_1 – selecção de 64 descritores vs MOLMAP completo.

	Concordância conjunto de treino de 1844 compostos	Concordância conjunto de validação de 49 compostos
Seleção 64 descritores (MOLMAP 29×29 1_0)	0,798	0,755
Seleção 64 descritores (MOLMAP 29×29 10_1)	0,827	0,816
841 descritores (MOLMAP 29×29 1_0 completo)	0,571	0,531
841 descritores (MOLMAP 29×29 10_1 completo)	0,577	0,327

Verifica-se que modelos baseados em selecções de 64 descritores proporcionam concordâncias mais elevadas do que no caso dos modelos treinados com a totalidade dos 841 descritores MOLMAP. Por um lado, MOLMAPs do tipo 1_0 permitem melhores performances do que tipo 10_1 quando são usados todos os 841 descritores. Por sua vez, quando são usadas selecções de 64 descritores MOLMAP, o tipo de MOLMAP 10_1 apresenta concordâncias mais elevadas tanto para o conjunto de treino como de validação.

Utilizando as características do melhor modelo (selecção de 64 descritores de MOLMAPs 29×29 do tipo 10_1), foram treinadas redes de *Kohonen* de outras dimensões para além de 20x20 (**Tabela III.35**).

Tabela III.35: Comparação, relativamente a valores de concordância, entre modelos baseados em redes de *Kohonen* de diferentes dimensões - modelos baseados em selecção de 64 descritores provenientes de MOLMAP 29×29 do tipo 10_1.

	Concordância conjunto de treino de 1844 compostos	Concordância conjunto de validação de 49 compostos
Rede <i>Kohonen</i> 10×10	0,802	0,796
Rede <i>Kohonen</i> 15×15	0,803	0,714
Rede <i>Kohonen</i> 20×20	0,827	0,816
Rede <i>Kohonen</i> 25×25	0,815	0,776
Rede <i>Kohonen</i> 29×29	0,835	0,714

É possível observar que o modelo baseado em rede de *Kohonen* 20×20 permite a obtenção de concordâncias mais elevadas tanto para o conjunto de treino como para o de validação. Foi assim seleccionado o modelo baseado na rede de *Kohonen* 20×20 treinada com a selecção de 64 descritores MOLMAP (29×29 do tipo 10_1). Este modelo foi usado para previsão do conjunto independente de teste de 49 compostos. Os resultados detalhados estão condensados na **Tabela III.36**.

Tabela III.36: Resultados obtidos, relativamente à concordância, especificidade e sensibilidade, para conjuntos de treino, validação e teste com melhor modelo baseado em rede neuronal de *Kohonen*.

Modelo rede de <i>Kohonen</i> 20×20 treinado com selecção de 64 descritores MOLMAP (29×29 10_1)			
	Conjunto de treino 1844 objectos	Conjunto de validação 49 objectos	Conjunto de teste 49 objectos
Concordância	0,827	0,816	0,796
Especificidade	0,879	0,913	0,920
Sensibilidade	0,774	0,731	0,666

Podem-se observar concordâncias relativamente elevadas para os conjuntos de treino, validação e teste. Para todos estes conjuntos observa-se que os valores relativos à sensibilidade são significativamente mais baixos que os respectivos valores de especificidade. Isto deve-se ao facto dos objectos mal classificados serem maioritariamente falsos negativos. Na **Figura III.35** estão mapeados os 49 compostos pertencentes ao conjunto independente de teste na rede de *Kohonen* treinada (e “colorida”) com o conjunto de treino.

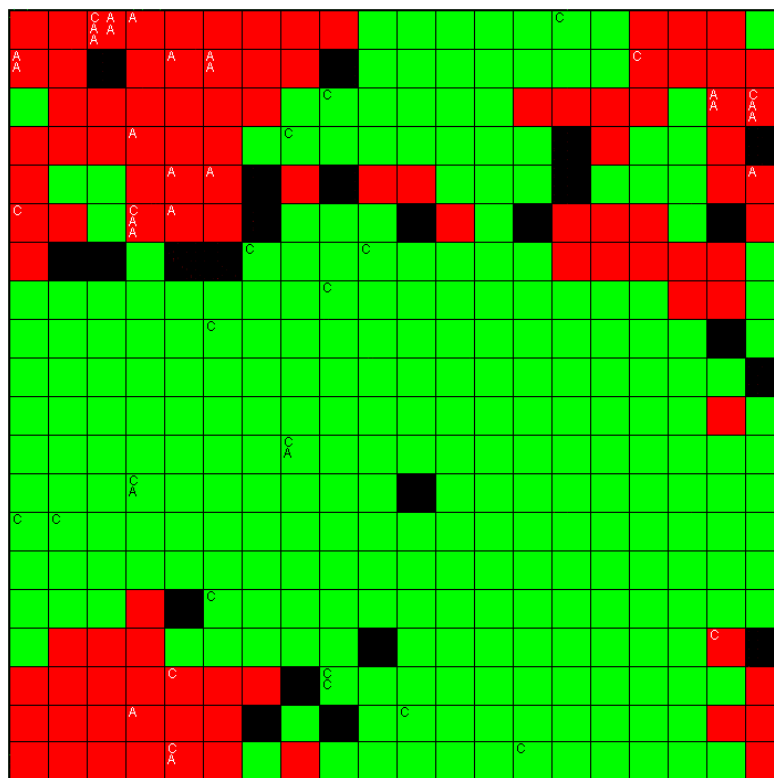


Figura III.35: Rede de *Kohonen* 20×20 treinada com selecção de 64 descritores MOLMAP (29×29 do tipo 10_1) onde foram mapeados os compostos pertencentes ao conjunto independente de teste. Os neurónios foram classificados de acordo com os objectos do treino que os activam: vermelho – não reage, verde – reage.

Usando a selecção de 64 descritores referentes a MOLMAPs 29×29 do tipo 10_1 foram treinados modelos baseados em algoritmo *Support Vector Machines*. A selecção do melhor modelo baseou-se em resultados obtidos para o conjunto de validação (**Tabela III.37**).

Tabela III.37: Valores de Concordância obtidos por modelos SVM variando o parâmetro C de complexidade – usando selecção de 64 descritores provenientes de MOLMAP 29×29 do tipo 10_1.

	Concordância conjunto de treino de 1844 compostos	Conjunto de treino, validação cruzada interna, 3 subconjuntos	Concordância conjunto de validação de 49 compostos
C = 1	0,882	0,819	0,837
C = 2	0,898	0,816	0,837
C = 4	0,909	0,822	0,837
C = 8	0,920	0,813	0,837
C = 16	0,926	0,793	0,796
C = 32	0,930	0,805	0,796
C = 64	0,934	0,798	0,776

É possível observar que todos os modelos construídos permitem a obtenção de concordâncias semelhantes para o conjunto de validação. Os melhores resultados foram obtidos com valores C de complexidade compreendidos entre 1 e 8. Prosseguiu-se o estudo com o modelo SVM baseado em C = 4 já que era o que apresentava valores de concordância mais robustos entre os diversos ensaios/conjuntos em análise. Para além do parâmetro C = 4 o modelo óptimo usa como função *kernel* a gaussiana com parâmetro sigma $\sigma = 0,126$. Este modelo assenta em 1057 *Support Vectors*.

Tabela III.38: Resultados obtidos, relativamente à concordância, especificidade e sensibilidade, para conjuntos de treino, validação e teste com modelo baseado em algoritmo *Support Vector Machines*.

	Modelo SVM, C = 4 treinado com selecção de 64 descritores MOLMAP (29×29 10_1)		
	Conjunto de treino 1844 objectos	Conjunto de validação 49 objectos	Conjunto de teste 49 objectos
Concordância	0,909	0,837	0,837
Especificidade	0,938	0,833	0,880
Sensibilidade	0,880	0,840	0,792

Tendo em conta os resultados apresentados na **Tabela III.38** é possível verificar que o modelo SVM permite a obtenção de concordâncias ligeiramente superiores em relação ao modelo baseado em rede de *Kohonen* optimizada. O modelo SVM não produz valores tão discrepantes entre especificidade e sensibilidade.

Este estudo permitiu a obtenção de modelos eficazes na previsão da reactividade de compostos na presença de butilamina, nucleófilo modelo. Adicionalmente verificou-se que MOLMAPs, baseados exclusivamente em propriedades físico-químicas das ligações, são eficazes na codificação de padrões de reactividade. No caso deste estudo os padrões de reactividade centram-se exclusivamente em propriedades físico-químicas de ligação. O método de selecção de descritores

baseado em mapeamento de ligações, classificadas segundo pertença a reagentes ou produtos, revelou-se eficaz para treino de rede de *Kohonen* e *Support Vector Machines*.

III-3.4 d): Estudo de Reactividade de Compostos na Presença de Cianoborohidreto de Sódio

O estudo contido nesta secção insere-se no mesmo contexto que o estudo anterior, ou seja, consiste no estudo da capacidade de descritores MOLMAP para codificar padrões de reactividade. Esses padrões baseiam-se em propriedades físico-químicas das ligações químicas, as mesmas 11 propriedades calculadas a partir do programa PETRA versão 4.0 que foram determinadas para o estudo **c)** anterior. A principal diferença entre o estudo **c)** anterior e o estudo desta secção é que agora, em vez de se estar a estudar a reactividade na presença de butilamina, está-se a estudar a reactividade na presença de cianoborohidreto de sódio. Para a geração de MOLMAPs foram usadas as mesmas redes auto-organizativas usadas no estudo **c)** anterior, como a representada na **Figura III.25**. Tal como no estudo **c)** anterior, a base de dados usada para este estudo consistiu exclusivamente em reacções envolvendo um reagente, o cianoborohidreto e um produto.



Esquema III.15: Representação genérica de uma reacção em que intervém o cianoborohidreto de sódio.

Nesta base de dados estão incluídos vários tipos de reacção, alguns dos quais aparecem representados na **Figura III.36**.

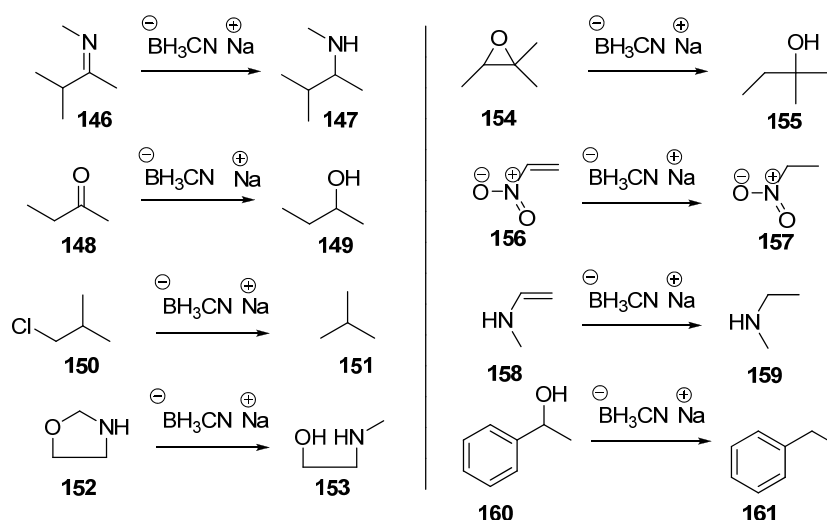


Figura III.36: Tipos de reacção presentes na base de dados reaccional em que intervém o cianoborohidreto de sódio.

Desta forma os reagentes presentes nesta base de dados vão ser usados como compostos reactivos. Por sua vez, para a obtenção de objectos que não vão reagir é necessária metodologia que codifique substruturas presentes em cada reagente que não vão ser alteradas durante cada reacção presente na base de dados original. Para a concretização deste objectivo foi seguido o mesmo procedimento utilizado no estudo **c)** anterior, ou seja, foi gerado MOLMAP para determinado reagente tal como foi gerado MOLMAP para o respectivo produto. Ambos os MOLMAP foram comparados (de acordo com Metodologia Computacional) de forma a identificar as ligações que não sofreram alteração na conversão de reagente em produto. Neste procedimento não foi contabilizado o MOLMAP do cianoborohidreto de sódio, visto este composto contribuir apenas com um equivalente de hidreto por equivalente de cianoborohidreto de sódio. Construiu-se assim um conjunto de treino baseado em 319 compostos que reagem, sendo incluídos neste conjunto 319 pseudocompostos que não reagem.

Usando o conjunto de treino foram construídos diferentes modelos *Random Forest* (1000 árvores, *mtry* = 25) baseados em descritores MOLMAP de diferentes dimensões: 29×29, 25×25, 20×20, 15×15, 10×10 e 7×7, e diferentes tipos: 1_0 e 10_1. Os diferentes modelos foram avaliados usando o conjunto de treino de 638 objectos e o conjunto de validação de 46 compostos (**Tabela III.39**).

Tabela III.39: Concordâncias relativamente a conjunto de treino, ensaio *Out of Bag* e conjunto de validação referentes a Modelos *Random Forest* baseados em diferentes tipos e dimensões de MOLMAP.

Dimensão do MOLMAP	MOLMAP tipo 10_1			MOLMAP tipo 1_0		
	Concordância			Concordância		
	Conjunto de treino	Conjunto de treino <i>Out of Bag</i>	Conjunto de validação	Conjunto de treino	Conjunto de treino <i>Out of Bag</i>	Conjunto de validação
29×29	1,000	0,917	0,913	0,995	0,879	0,935
25×25	1,000	0,922	0,891	1,000	0,876	0,913
20×20	1,000	0,926	0,826	1,000	0,867	0,891
15×15	1,000	0,964	0,804	1,000	0,890	0,957
10×10	1,000	0,978	0,652	1,000	0,892	0,848
7×7	1,000	0,988	0,609	1,000	0,917	0,826

Todos os modelos construídos foram avaliados na capacidade de previsão, sendo realizado ensaio *Out of Bag* com conjunto de treino, assim como também foi utilizado conjunto independente de validação de 46 compostos. Tendo em atenção esses dois itens pode-se observar que MOLMAPs de dimensões inferiores a 15×15 apresentam elevadas concordâncias para ensaios *Out of Bag* mas baixas concordâncias para o conjunto de validação. Este facto é particularmente evidente em MOLMAPs do tipo 10_1. O motivo para este tipo de resultados foi abordado no estudo c) anterior. Foram assim seleccionados MOLMAPs de dimensões 15×15 do tipo 1_0 como os mais indicados para previsão de classe reaccional já que permitiram a obtenção de concordância mais elevada para o conjunto de validação (0,957). Estes MOLMAPs foram utilizados posteriormente na optimização do valor de *mtry* sendo obtidos os resultados da Tabela III.40.

Tabela III.40: Valores de Concordância obtidos com diferentes valores de *mtry* de forma a otimizar modelo de previsão de reactividade na presença de cianoborohidreto de sódio. Conjunto de treino *Out of Bag* e conjunto de validação.

Valor de <i>mtry</i>	MOLMAP tipo 1_0 dimensão 15×15	
	Concordância	
	Conjunto de treino <i>Out of Bag</i>	Conjunto de validação
15	0,890	0,957
25	0,890	0,957
50	0,890	0,957
75	0,900	1,000
100	0,900	1,000
150	0,897	0,978
200	0,897	0,978
225	0,898	0,978

Pode-se constatar que os modelos que apresentam os melhores resultados para o conjunto de validação baseiam-se em valores de $mtry = 75$ e 100, sendo seleccionado $mtry = 75$ já que o respectivo modelo RF foi gerado mais rapidamente. O valor de $mtry$ seleccionado permite a obtenção de concordância máxima para o conjunto de validação (1,000). O modelo RF seleccionado foi usado na previsão de reactividade para o conjunto independente de teste. Resultados completos relativamente a este conjunto, conjuntos de treino e validação aparecem condensados na **Tabela III.41**.

Tabela III.41: Resultados obtidos, relativamente à concordância, especificidade e sensibilidade por modelo RF seleccionado para conjuntos de treino, validação e teste.

Modelo <i>Random Forest</i> 1000 árvores, $mtry = 75$ treinado com MOLMAPs dimensão 15×15 tipo 1_0		
	Conjunto de validação 46 objectos	Conjunto de teste 48 objectos
Concordância	1,000	0,938
Especificidade	1,000	1,000
Sensibilidade	1,000	0,880

É possível observar que modelo seleccionado permite a obtenção de valores elevados de concordância para o conjunto independente de teste (concordância = 0,938). Houve apenas três objectos mal classificados, mais precisamente três falsos negativos. As três previsões incorrectas correspondem a probabilidades de reacção acima 0,400. Os compostos do conjunto de treino mais próximos a estes três compostos são pseudocompostos que não reagem. Em dois casos, reacção de grupo hidroxilo não é detectada. Isto deve-se ao facto dos respectivos objectos do conjunto de treino mais próximos se basearem em reacções em que, para além do desaparecimento do grupo hidroxilo, também ocorre o desaparecimento de olefina. Por sua vez a funcionalidade olefina não está presente nestes compostos do conjunto de teste. O terceiro objecto é mal classificado já que possui combinação de funcionalidades não existentes em objectos do conjunto de treino, sendo pouco próximo a qualquer objecto desse conjunto.

Ainda relativamente ao conjunto independente de teste foi gerada uma curva ROC de maneira a assegurar que o incremento na fracção de positivos verdadeiros não está relacionado com o incremento do número total de previsões positivas independentemente de serem verdadeiras ou falsas (**Figura III.37**). Esta medida de não correlação é a área entre a curva ROC e a linha de não discriminação. Foi calculado um valor de 0,991.

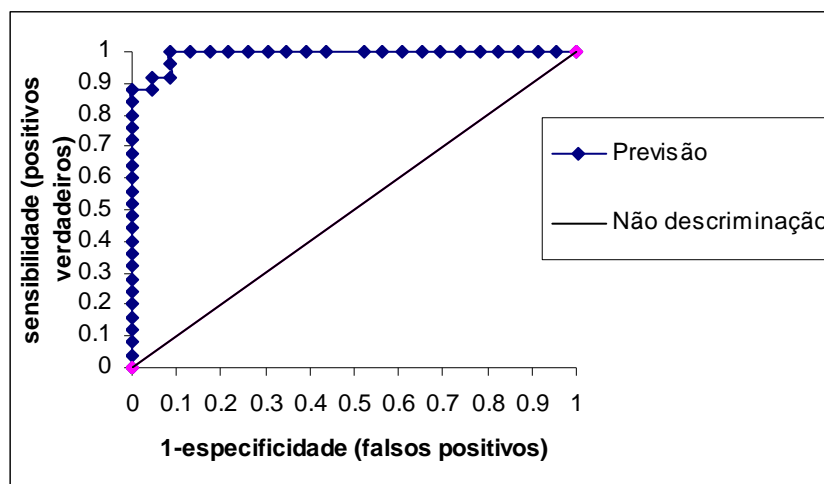


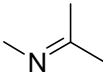

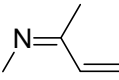

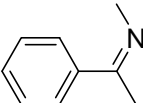

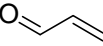

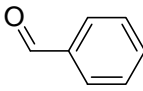

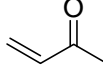
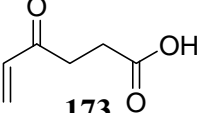
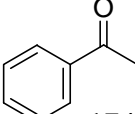
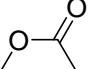
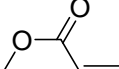
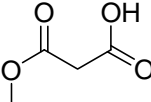
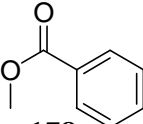
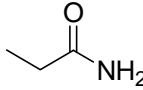
Figura III.37: Curva ROC obtida com os resultados do melhor modelo RF para os 48 compostos do conjunto independente de teste.

Verificou-se também a presença de três objectos no conjunto de validação e três objectos no conjunto de teste com o mesmo MOLMAP relativamente a objectos do conjunto de treino. Ainda assim as moléculas em causa não são as mesmas (nem todas as ligações de uma molécula são contabilizadas em MOLMAP – ver Metodologia Computacional).

O melhor modelo RF foi adicionalmente validado dispondo aleatoriamente a classe reaccional dos objectos do conjunto de treino. Esse conjunto depois de modificado foi usado para treinar RF com as mesmas características. Foi obtida concordância = 0,530 em ensaio *Out of Bag*, 0,520 e 0,460 para conjuntos de validação e teste respectivamente.

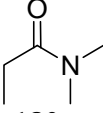
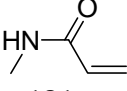
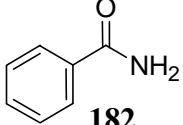
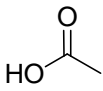
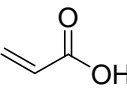
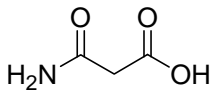
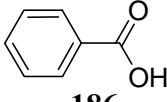
O melhor modelo RF foi ainda validado com um conjunto de 25 compostos construídos a partir de informação de quimioselectividade descrita na literatura.⁴⁵ As estruturas aí presentes são simples e incluem diferentes grupos funcionais, de forma a evidenciar a capacidade do modelo em prever a quimioselectividade do cianoborohidreto de sódio (**Tabela III.42**). A concordância para este conjunto é máxima (1,000). Os resultados demonstram que o modelo RF, baseado exclusivamente em descritores físico-químicos, consegue distinguir as funcionalidades reactivas (iminas, cetonas e aldeídos) das funcionalidades não reactivas (ésteres, ácidos carboxílicos e amidas).

Tabela III.42-1: Lista de compostos gerados a partir de informação recolhida na literatura relativamente à reactividade de determinadas funcionalidades na presença de cianoborohidreto de sódio.

 162	 163	 164
c (literatura)	c (literatura)	c (literatura)
c (previsão)	c (previsão)	c (previsão)
 165	 166	 167
c (literatura)	c (literatura)	c (literatura)
c (previsão)	c (previsão)	c (previsão)
 168	 169	 170
c (literatura)	c (literatura)	c (literatura)
c (previsão)	c (previsão)	c (previsão)
 171	 172	 173
c (literatura)	c (literatura)	c (literatura)
c (previsão)	c (previsão)	c (previsão)
 174	 175	 176
c (literatura)	a (literatura)	a (literatura)
c (previsão)	a (previsão)	a (previsão)
 177	 178	 179
a (literatura)	a (literatura)	a (literatura)
a (previsão)	a (previsão)	a (previsão)

Aparece incluída a classe reactiva baseada na literatura e classe prevista pelo modelo: a – não reage, c – reage.

Tabela III.42-2: Continuação da lista de compostos gerados a partir de informação recolhida na literatura relativamente à reactividade na presença de cianoborohidreto de sódio.

 180	 181	 182
a (literatura)	a (literatura)	a (literatura)
a (previsão)	a (previsão)	a (previsão)
 183	 184	 185
a (literatura)	a (literatura)	a (literatura)
a (previsão)	a (previsão)	a (previsão)
 186		
a (literatura)		
a (previsão)		

Tendo em conta os resultados evidenciados para os diversos conjuntos de compostos pode-se afirmar que o melhor modelo RF é um modelo que tem muito boa capacidade de previsão para novos casos. Outro aspecto tão importante como a capacidade de previsão é a interpretabilidade de determinado modelo relativamente às variáveis mais importantes a modelar determinada propriedade. Para responder a esta questão foi construído um modelo Árvore de Decisão baseado em MOLMAPs 15×15 do tipo 1_0 (**Figura III.38**).

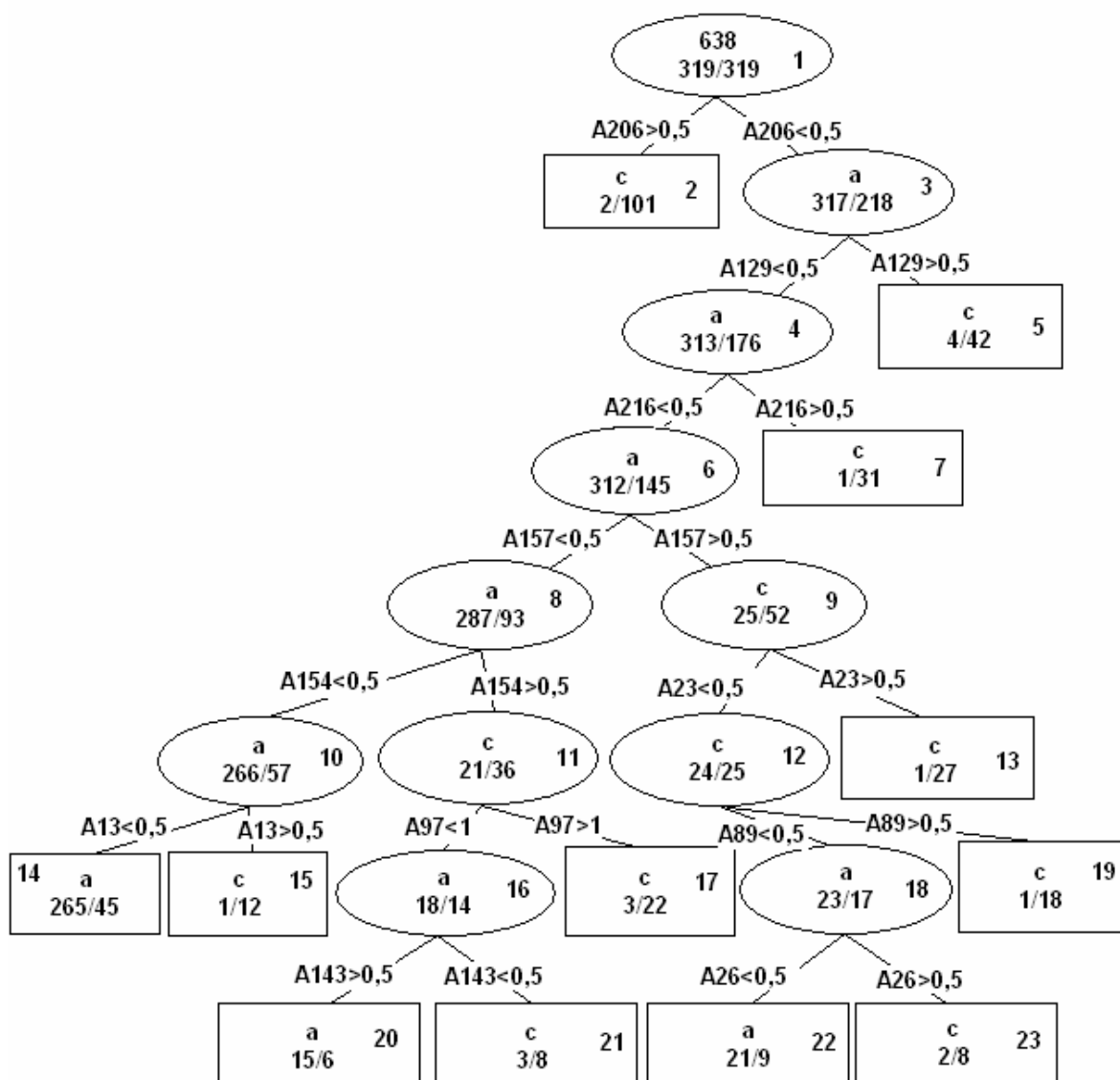


Figura III.38: Esquema de árvore de decisão, construída com MOLMAPs 15x15 do tipo 1_0 usada para modelar a reactividade de composto na presença de cianoborohidreto de sódio. Classe c – reage, classe a – não reage.

Através da análise realizada a esta árvore de decisão foi possível identificar regras que estão de facto associadas a tipos de ligações reactivas na presença de cianoborohidreto de sódio. As únicas duas excepções são os descritores A143 e A157 que representam ligações carbono-carbono aromáticas. Encontra-se representada na **Tabela III.43** uma análise detalhada dos descritores seleccionados por árvore de decisão. Esta Árvore de Decisão permitiu a obtenção de concordância = 0,850 para conjunto independente de teste.

Tabela III.43: Tipos de ligação representadas em árvore de decisão da Figura III.38.

Descritor	Tipo de ligação representada
A206	C=N de iminas ligação C=S em funcionalidade éster ditioalquílico
A129	C-H olefinica em composto carbonílico α,β - insaturado
A216	C=C em composto α,β – insaturado e 1-Bromoalcenos
A157	C=C aromáticas e C-C simples em que um dos carbonos é aromático (artifício relacionado com a distribuição de compostos do conjunto de treino)
A154	C-H em que carbono pertence a funcionalidade aldeído / C-C em que carbono pertence a funcionalidade cetona
A23	C-N simples de funcionalidade enamina
A89	Carbonilo pertencente a funcionalidade cetona
A26	C-O de epóxido
A13	C-N associado a enamina
A97	C-H em que carbono pertence a grupo electroatractor como é o caso de carbonilos e nitrilos
A143	C=C aromáticas (artifício relacionado com a distribuição de compostos do conjunto de treino)

Tendo em conta os descritores representados na **Tabela III.43** é possível observar que a grande maioria dos tipos de ligação seleccionados são efectivamente ligações reactivas na presença de cianoborohidreto de sódio, desde ligações C=N e C=S de imina e éster ditioalquílico (A206), ligações C-H associadas a aldeídos, compostos carbonílicos α,β – insaturados e a grupos electro-atractores como carbonilos e nitrilos (A154, A129 e A97 respectivamente), ligações C-N de enamina (A23 e A13) tal como ligações C-O de epóxidos (A26) e carbonilos de cetona (A89).

A discussão sobre esta árvore de decisão pode ajudar na interpretação dos descritores definidos como os mais importantes segundo o melhor modelo *Random Forest*, até porque alguns descritores são os mesmos que foram seleccionados pela árvore de decisão. Os dez descritores mais importantes segundo o melhor modelo RF aparecem representados na **Tabela III.44**.

Tabela III.44: Os dez descritores mais importantes segundo melhor modelo RF – 1000 árvores, $mtry = 75$ obtido com MOLMAPs 15x15 do tipo 1_0 e respectivos valores de importância.

Posição	Descritor	Importância (ver Capítulo I)
1	A206	1,891
2	A216	1,708
3	A112	1,441
4	A129	1,429
5	A98	1,394
6	A139	1,297
7	A86	1,274
8	A88	1,273
9	A13	1,248
10	A154	1,244

Os dez descritores mais importantes seleccionados pelo melhor modelo *Random Forest* estão na **Tabela III.44**. Pode-se observar que cinco dos descritores mais importantes por *Random Forest* são também utilizados pela árvore de decisão representada na **Figura III.38**. Concretamente, o descritor A206 colocado na primeira posição, o descritor A216 colocado na segunda posição, o descritor A129 colocado na quarta posição, o descritor A13 colocado na nona posição e finalmente o descritor A154 colocado na décima posição. Estes descritores, pela análise à árvore de decisão, representam ligações reactivas na presença de cianoborohidreto de sódio.

O descritor A112 representa ligações carbono-hidrogénio cujo carbono está ligado a carbonilo de cetona ou a olefina de carbonilo α,β -insaturado. O descritor A98 representa ligações carbono-hidrogénio em anel aromático. Este tipo de ligação está colocada em posição *orto* em relação a grupo electroattractor, por exemplo aldeídos. Na sexta posição está o descritor A139 que representa ligações carbono-carbono e carbono-hidrogénio associadas às funcionalidades representadas na **Figura III.39**

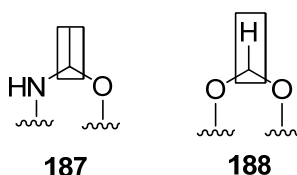


Figura III.39: Funcionalidades reactivas na presença de cianoborohidreto de sódio.

Estas duas funcionalidades são reactivas segundo exemplos da base de dados original em que se baseou este estudo. Por sua vez, na sétima posição surge o descritor A86, associado a grupo hidroxilo (em posição benzílica e não benzílica) que na base de dados reaccional reage quando se

apresenta em posição bezilica. Finalmente, o descritor A88 representa carbonilo de aldeído ou cetona α,β -insaturados.

Desta forma foi obtido modelo RF com boa capacidade de previsão na reactividade na presença de cianoborohidreto de sódio, capaz de distinguir funcionalidades reactivas de não reactivas, de acordo com o conhecimento existente numa base de dados de reacções.

III-4: Conclusão

No estudo **a)**, as experiências para a previsão do mecanismo reaccional dado o composto a reagir com butilamina permitiram obter previsões acertadas para um máximo de 76% do conjunto independente de teste. O melhor modelo baseou-se em algoritmo RF e descritores MOLMAP baseados em propriedades topológicas e físico-químicas. Também foi obtido um modelo baseado em rede de *Kohonen* treinada por descritores MOLMAP seleccionados por algoritmos genéticos que revelou um agrupamento satisfatório das 6 diferentes classes reaccionais e 69% de previsões certas para o conjunto de teste. As limitações destes modelos estão relacionadas com algumas incapacidades de discriminação de ligações relevantes por parte dos mapas auto-organizativos que geram os MOLMAP, com a presença de múltiplas funcionalidades potencialmente reactivas (segundo mecanismos diferentes) na mesma molécula e ainda com funcionalidades pouco representadas no conjunto de treino.

Compreensivelmente, os modelos treinados no estudo **b)**, em que as reacções foram classificadas quanto aos mecanismos reaccionais a partir dos MOLMAPs dos reagentes *e dos produtos*, apresentaram percentagens de acerto mais elevadas em conjuntos de teste – até 91%.

Nos estudos **a)** e **b)**, o emprego de conjunto misto de propriedades topológicas e físico-químicas para representar as ligações químicas foi mais eficaz na codificação em MOLMAPs de padrões associados aos diferentes mecanismos reaccionais em estudo.

Nos estudos **c)** e **d)**, as experiências de construção de modelos para prever se um composto reage ou não na presença de um dado reagente (butilamina ou NaCNBH₃) revelaram que é possível retirar informação sobre a reactividade de compostos implicitamente contida em bases de dados de reacções orgânicas, com base exclusivamente em propriedades físico-químicas empíricas calculadas para átomos e moléculas.

Os descritores moleculares MOLMAP permitiram derivar informação negativa acerca da reactividade frente aos dois reagents específicos a partir das bases de dados de reacções. Combinando esta informação com MOLMAPs de compostos reactivos o algoritmo *Random Forest* foi capaz de extrair conhecimento acerca da reactividade que foi posteriormente validado com conjuntos de previsão independentes. A informação negativa derivada de MOLMAPs consiste em ligações que não sofrem alteração durante a reacção. Uma desvantagem deste método, que deduz informação sobre ligações não-reactivas a partir de reacções que ocorrem, é a inclusão de falsos negativos quando os reagentes possuem mais do que um local possível de reacção.

O método só consegue extrair informação negativa se a reacção causar nas ligações envolvidas alteração das propriedades usadas. Para além disso as reacções devem ser apresentadas o mais possível acertadas em termos estequiométricos.

Apesar dos bons resultados obtidos com reacções da bases de dados de BuNH₂, o modelo não pôde ser aplicado com sucesso à previsão de reactividade com péptido de lisina. Os dados usados relativos a ensaios de reactividade com péptidos referem-se a condições muito específicas, nomeadamente concentrações, solvente e temperatura. Ao invés, a base de dados usada para treinar o modelo incluiu reacções de origens muito diversas, realizadas em condições que não foram homogeneizadas. Este facto explica muitas previsões que estão de acordo com reacções observadas na base de dados mas que não ocorrem no teste com o péptido (por exemplo reacção de ácidos carboxílicos). A utilização de condições muito específicas torna também relevantes propriedades globais da molécula (como seja a solubilidade no solvente usado) que não podem ser aproximadas pelos descritores MOLMAP usados. Modelos mais realistas poderão ser construídos com dados obtidos em condições bem padronizadas à medida que estes estejam disponíveis em quantidade suficiente.

Os quatro estudos aqui contidos representam contribuições quimio-informáticas complementares no estudo de reacções químicas, abrangendo classificação de reacções em termos do mecanismo se forem conhecidos os reagentes e produtos, previsão do mecanismo dada a estrutura do reagente e previsão sobre a possibilidade ou não de reacção dado um composto.

II-5: Bibliografia:

- (1) Aptula, A. O.; Patlewicz, G.; Roberts, D. W.; *Chem. Res. Toxicol.*, **2005**, 18, 1420-1426.
- (2) Directiva 2003/15/EC do Parlamento Europeu de 27 Fevereiro 2003 em emenda à Directiva 76/768/EEC. OJ L066, 26-35, 11 Março de 2003.
- (3) Gerberick, G. F.; Vassallo, J. D.; Bailey, R. E.; Chaney, J. G.; Morrall, S. W.; Lepoittevin, J-P.; *Toxicol. Sci.*, **2004**, 81, 332-343.
- (4) Zhang, Q.; Aires-de-Sousa, J.; *J. Chem. Inf. Model.*, **2005**, 45(6), 1775-1783.
- (5) Latino, D. A. R. S.; Aires-de-Sousa, J.; *Angew. Chem. Int. Ed.*, **2006**, 45, 2066-2069.
- (6) Gupta, S.; Matthew, S.; Abreu, P. M.; Aires-de-Sousa, J.; *Bioorg. Med. Chem.*, **2006**, 14, 1199-1206.
- (7) Zhang, Q.; Aires-de-Sousa, J.; *J. Chem. Inf. Model.*, **2007**, 47, 1-8.
- (8) Todd, M. H.; *Chem. Soc. Rev.*, **2005**, 34, 247-266.
- (9) Simon, V.; Gasteiger, J.; Zupan, J.; *J. Am. Chem. Soc.*, **1993**, 115, 9148-9159.
- (10) Satoh, H.; Funatsu, K.; *J. Chem. Inf. Comput. Sci.*, **1996**, 36, 173-184.
- (11) Socorro, I. M.; Taylor, K.; Goodman, J. M.; *Organic Letters*, **2005**, 7, 3541-3544.
- (12) Kulkarni, S. A.; Zhu, J.; Blechinger, S.; *Xenobiotica*, **2005**, 35, 955-973.
- (13) Ridder, L.; Wagener, M.; *ChemMedChem*, **2008**, 3, 821-832.
- (14) Oh, M.; Yamada, T.; Hattori, Masahiro, Goto, S.; Kanehisa, M.; *J. Chem. Inf. Model.*, **2007**, 47, 1702-1712.
- (15) Boyer, S.; Arnby, C. H.; Carlsson, L.; Smith, J.; Stein, V.; Glen, R. C.; *J. Chem. Inf. Model.*, **2007**, 47, 583-590.
- (16) Kulkarni, S. A.; Moir, D.; Zhu, J.; *SAR and QSAR in Environ. Res.*, **2007**, 5-6, 459-514.
- (17) Mekenyan, O.; Todorov, M.; Serafimova, R.; Stoeva, S.; Aptula, A.; Finking, R.; Jacob, E.; *Chem. Res. Toxicol.*, **2007**, 20, 1927-1941.
- (18) Huth, J. R.; Mendoza, R.; Olejniczak, E. T.; Johnson, R. W.; Cothron, D. A.; Liu, Y.; Lerner, C. G.; Chen, J.; Hajduk, P. J.; *J. Am. Chem. Soc.*, **2005**, 127, 217-224.
- (19) Huth, J. R.; Song, D.; Mendoza, R. R.; Black-Schaefer, C. L.; Mack, J. C.; Dorwin, S. A.; Lador, U. S.; Severin, J. M.; Walter, K. A.; Bartley, D. M.; Hajduk, P. J.; *Chem. Res. Toxicol.*, **2007**, 20, 1752-1759.
- (20) Metz, J. T.; Huth, J. R.; Hajduk, P. J.; *J. Comput. Aided Mol. Des.*, **2007**, 21, 139-144.
- (21) Huang, D.; Ou, B.; Prior, R. L.; *J. Agric. Food Chem.*, **2005**, 53, 1841-1856.
- (22) Pease, C. K. S.; *Toxicology*, **2003**, 192, 1-22.
- (23) Momma, J.; Kitajima, S.; Inoue, T.; *Toxicology*, **1998**, 126, 75-82.
- (24) Gerberick, G. F.; Ryan, C. A.; Kern, P. S.; Dearman, R. J.; Kimber, I.; Patlewicz, G. Y.; Basketter, D. A.; *Contact Dermatitis*, **2004**, 50, 274-288.
- (25) Directiva 429 para teste de compostos químicos da OCDE, em <http://www.oecd.org/>
- (26) Roberts, D. W.; Aptula, A. O.; Patlewicz, G.; *Chem. Res. Toxicol.*, **2007**, 20, 44-60.
- (27) Ashby, J.; Basketter, D. A.; Paton, D.; Kimber, I.; *Toxicology*, **1995**, 103, 177-194.
- (28) Patlewicz, G.; Roberts, D. W.; Uriarte, E.; *Chem. Res. Toxicol.*, **2008**, 21, 521-541.
- (29) Roberts, D. W.; *Chem. Res. Toxicol.*, **1995**, 8, 545-551.
- (30) Patlewicz, G. Y.; Basketter, D. A.; Pease, C. K. S.; Wilson, K.; Wright, Z. M.; Roberts, D. W.; Bernard, G.; Arnau, E. G.; Lepoittevin, J-P.; *Contact Dermatitis*, **2004**, 50 (2), 91-97.
- (31) Li, S.; Fedorowicz, A.; Singh, H.; Soderholm, S. C.; *J. Chem. Inf. Model.*, **2005**, 45, 952-964.
- (32) Landsteiner, K.; Jacobs, J. L.; *J. Exp. Med.*, **1936**, 64, 625-639., citado em: Ashby, J.; Basketter, D. A.; Paton, D.; Kimber, I.; *Toxicology*, **1995**, 103, 177-194.
- (33) Aptula, A. O.; Patlewicz, G.; Roberts, D. W.; Schultz, T. W.; *Toxicol. in Vitro*, **2006**, 20, 239-247.
- (34) Natsch, A.; Emter, R.; Ellis, G.; *Toxicol. Sci.*, **2009**, 107, 106-121.
- (35) Gerberick, G. F.; Vassallo, J. D.; Foertsch, L. M.; Price, B. B.; Chaney, J. G.; Lepoittevin, J-P.; *Toxicol. Sci.*, **2007**, 97, 417-427.

- (36) Almeida, D. I. C.; Dissertação de Mestrado “*Estudos sobre reactividade química da glutathione com métodos de aprendizagem automática*”, UNL, Lisboa, **2008**.
- (37) Satoh, H.; Funatsu, K.; Takano, K.; Nakata, T.; *Bull. Chem. Soc. Jpn.*, **2000**, 73, 1955-1965.
- (38) Consultar página web: <http://infochem.de/>.
- (39) Download e informação adicional sobre programas Chemaxon em <http://www.chemaxon.com/>
- (40) O Programa PETRA pode ser testado na página web <http://www2.chemie.uni-erlangen.de> sendo desenvolvido pela empresa Molecular Networks GmbH, <http://www.mol-net.de>
- (41) Aires-de-Sousa, J.; *Chemom. Intell. Lab. Syst.*, **2002**, 61, 167-173.
- (42) Os applets JATOON estão disponíveis em <http://www.dq.fct.unl.pt/staff/jas/jatoon>.
- (43) Ambiente estatístico R disponível para download na página web:
<http://www.r-project.org/>
- (44) Enoch, S. J.; Cronin, M. T. D.; Schultz, T. W.; Madden, J. C.; *Chem. Res. Toxicol.*, **2008** 21, 513-520.
- (45) Clayden, J.; Greeves, N.; Warren, S.; Wothers, P.; *Organic Chemistry*, Oxford University Press, Oxford, **2001**.

Capítulo IV

Glossário

Ames, teste: Teste padrão para determinar mutagenicidade de um composto químico, realizado na presença de *Salmonella typhimurium*.

Análise de Componentes Principais (PCA): É um método que visa a redução de variáveis originais utilizadas num estudo, num conjunto menor de novas variáveis (PCA's ou variáveis latentes), função das originais, explicando da mesma forma a variância nas variáveis originais. Estas novas variáveis, ao contrário das primeiras, são ortogonais, ou seja não estão correlacionadas. De uma forma sequencial, partindo das m variáveis dos n objectos e representando-as num sistema de eixos m -dimensional como pontos, o método procura a linha (combinação linear das variáveis originais) que melhor se aproxime desses pontos, sendo esse o primeiro componente principal PC1 e que representa a variância máxima no conjunto de variáveis originais. O segundo componente principal (PC2) vai ser ortogonal (perpendicular) ao primeiro e representa a segunda maior variância, sendo menos importante que o primeiro; e assim sucessivamente irão aparecer novos componentes principais. Estas novas variáveis vão estar representadas por um *score*, ou seja; a sua coordenada e pelo *loading*, a forma como as variáveis originais se combinam para obter o *score*. Desta forma a matriz das variáveis originais (n objectos e m variáveis) \mathbf{X} , é função da matriz \mathbf{T} dos *scores* (n objectos, e d novas variáveis) e da matriz \mathbf{P} dos *loadings* (d novas variáveis e m variáveis): $\mathbf{X} = \mathbf{TP}^T$.

Anisotropia: Associada a um composto, é a propriedade de ser dependente da direcção, em oposição à isotropia que significa homogeneidade em todas as direcções. Essa heterogenia pode-se reflectir na diferença observada em determinadas propriedades segundo se escolha uma ou outra direcção para realizar a medição.

Antigénio: É uma substância que estimula a geração de anticorpos, causando uma resposta imune. Algumas proteínas e polisacáridos são substâncias antigénicas.

Bayesiano, método de classificação: Modelo automático de aprendizagem de problemas de classificação baseados na probabilidade condicional de uma determinada classe C dadas as variáveis de input F_1, F_2, \dots, F_n

$$p(C|F_1, F_2, \dots, F_n) = \frac{p(C).p(F_1, F_2, \dots, F_n|C)}{p(F_1, F_2, \dots, F_n)}$$

Nesta equação, o valor do denominador vai ser constante daí que seja relevante centrar-nos no numerador dessa equação:

$$p(C).p(F_1, F_2, \dots, F_n|C)$$

Que é equivalente a:

$$p(C).p(F_1, F_2, \dots, F_n | C) = p(C).p(F_1 | C).p(F_2, \dots, F_n | C, F_1) = p(C).p(F_1 | C).p(F_2 | C, F_1).p(F_3, \dots, F_n | C, F_1, F_2) = \dots$$

O modelo Bayesiano simplifica esta equação, assumindo uma independência condicional entre todas as variáveis de *input* F_1, F_2, \dots, F_n em que F_i é condicionalmente independente de outras variáveis de *input* $F_j (j \neq i)$. Nessas condições a equação de cima fica:

$$p(C).p(F_1, F_2, \dots, F_n | C) = p(C).p(F_1 | C).p(F_2 | C) \dots p(F_n | C)$$

È por assumir esta inter-relação independente entre as variáveis, que este algoritmo é também denominado algoritmo naive Bayes.

Célula unitária: Unidade fundamental que, quando repetida ou submetida a operações de simetria, irá constituir uma rede cristalina.

Classificação, modelo: Modelo genérico que se baseia na relação entre as variáveis de *input* (independentes) e as variáveis dependentes, estas últimas são classes, onde podem ser agrupados objectos com parâmetros de *input* diferentes.

Coefficiente de expansão térmica, (α_p): Tal como o próprio nome indica é uma medida que permite quantificar a alteração do espaço que uma substância sofre com a mudança de temperatura. É definida matematicamente por $\alpha_p = -\frac{1}{\rho} \left(\frac{\partial \rho}{\partial T} \right)_p$, em que ρ é a densidade e T a temperatura. Tem como unidades o K^{-1} .

Compressibilidade isotérmica, (k_T): Como o próprio nome indica é uma medida relativa da mudança de volume duma substância com a variação da pressão. Está matematicamente definido por $k_T = \frac{1}{\rho} \left(\frac{\partial \rho}{\partial P} \right)_T$, em que ρ é a densidade e P a pressão. Tem como unidades o Pa^{-1} .

Conjunto de treino: Conjunto de objectos utilizados para construir um modelo de aprendizagem.

Conjunto de teste: Conjunto de objectos que não foram utilizados para treinar um modelo, e que são previstos depois do treino para validar o modelo.

Constante dieléctrica, (ϵ): Ou constante dieléctrica estática, é o rácio entre a quantidade de energia eléctrica armazenada quando um potencial é aplicado ϵ_s e a permissividade do vácuo ϵ_0 . A constante

dieléctrica é definida matematicamente como: $\epsilon = \frac{\epsilon_s}{\epsilon_0}$.

Cristal líquido: Composto que mesmo estando no estado líquido, irá ter ordem no arranjo entre as moléculas, tal como se de um sólido cristalino se tratasse.

Descritores 0D: Representam características globais da molécula tal como o peso molecular, ou o número total de átomos de uma molécula.

Descritores 1D: Descritores que representam alguma organização estrutural da molécula, tal como a contagem determinadas funcionalidades.

Descritores 2D: Descritores que representam a conectividade de determinada molécula.

Descritores 3D: Descritores que representam a estrutura tridimensional da molécula.

Descritores 4D: Descritores que codificam a superfície molecular associada à estrutura tridimensional de uma dada molécula.

Energia de Gibbs de fusão, (ΔG_{fus}): Definido como a quantidade máxima de trabalho não expansivo que pode ser retirado de um sistema fechado. Na prática é uma medida de espontaneidade de ocorrência de um determinado fenómeno, neste caso a fusão, estando matematicamente definido por:

$$\Delta G_{fus} = \Delta H_{fus} - T\Delta S.$$

Entalpia de fusão, (ΔH): É o calor trocado pelo sistema durante o processo de fusão, denominado entalpia por se processar a pressão constante.

Entropia de fusão, (ΔS): É o parâmetro que é definido por: $\Delta S = \int \frac{\partial q_{rev}}{T}$, em que ∂q_{rev} é a variação infinitesimal de calor e T é a temperatura de fusão.

Erro médio quadrático, (MSE): medida dos desvios entre valores experimentais de cada objecto e_i e

respectivos valores previstos p_i , definida por: $MSE = \frac{\sum_{i=1}^N (e_i - p_i)^2}{N}$, em que N é o número total de objectos.

Esmética, fase: Fase líquida cristalina, em que as moléculas têm ordem posicional segundo uma determinada direcção.

Euclidiana, Distância: Matematicamente, a distância Euclideana é a distância entre dois pontos determinados (x_1, x_2, \dots, x_m) e (w_1, w_2, \dots, w_m) , definidos num espaço m -dimensional, sendo determinada pela seguinte fórmula:

$$distância_Euclideana = \sqrt{\sum_{i=1}^m (x_i - w_i)^2}$$

Folding: É o processo físico-químico através do qual um polipéptido se organiza tridimensionalmente de forma a apresentar funcionalidade e propriedades características.

Heurístico: Método simples para resolver um determinado problema, método rápido e normalmente informal.

Índice de refração, (R_I): Uma medida que define o rácio $\frac{c_0}{c}$, em que c é a velocidade da luz no meio em questão e c_0 a velocidade da luz no vácuo.

Interacções, n- π : Interações intermoleculares não covalentes, que se baseiam na interacção entre electrões não ligantes e orbitais π .

Interacções, π - π : São interações intermoleculares não covalentes, que se baseiam na sobreposição de orbitais p de sistemas π de diferentes moléculas.

Iónica, forças: Forças de interacção entre iões originadas pelas suas cargas.

Isotrópica, fase: Fase líquida, em que as moléculas não possuem nenhuma ordem entre si.

Leave one out, validação: É um tipo de validação em que partindo de um conjunto de n objectos, $n-1$ vão ser usados para treinar um modelo de aprendizagem automática, enquanto que o objecto de fora é utilizado para validar o modelo, ou seja, obter previsão. Este procedimento é repetido $n-1$ vezes adicionais para que todos os objectos tenham ficado uma vez de fora do treino.

logP_{ow}: É a proporção das concentrações de um composto entre uma fase aquosa e uma fase de octanol, e que mede a hidrofilicidade ou a hidrofobicidade de um composto. Aparece definido como:

$$\log P_{ow} = \log \left(\frac{[soluto]_{oc\ tan\ ol}}{[soluto]_{agua}} \right)$$

London, forças dispersivas: Classe particular de forças de Van der Waals. Baseiam-se em interações intermoleculares fracas, provocadas pela formação de dipolos instantâneos.

Mesofase: Cada uma das fases presentes no cristal líquido.

Metabolito: É um composto, produto ou intermediário de metabolismo, estando geralmente associado a compostos de pequenas dimensões. Existem dois tipos de metabolitos distintos, por um lado há os denominados metabolitos primários, responsáveis por funções fundamentais num determinado organismo como é o caso do seu crescimento e reprodução. Por outro lado há também os metabolitos secundários que não estando envolvidos em nenhuma função fundamental à manutenção da espécie, podem ter valor ecológico relevante ora como antibióticos, pigmentos entre outros.

Não supervisionado, método: É um tipo de aprendizagem automática, mas que ao contrário do método supervisionado, não estabelece uma relação entre parâmetros de *input* e parâmetros de *output*. Este método olha exclusivamente para os parâmetros de *input* para a construção de um modelo.

Needleman-Wunsch, Algoritmo: Algoritmo proposto em 1970 por Saul Needleman e Christian Wunsch, capaz de realizar um alinhamento global entre duas sequências. É usado habitualmente em bio-informática para alinhar sequências de nucleótidos, tal como sequências de proteínas. Este algoritmo baseia-se no emprego de uma matriz de similitude entre as

- A G C T

A 10 -1 -3 -4

G -1 7 -5 -3

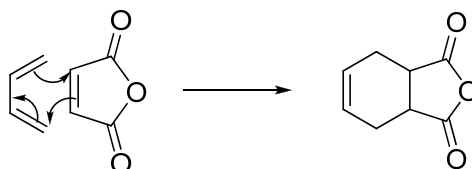
C -3 -5 9 0

T -4 -3 0 8

diferentes unidades fundamentais das sequências para determinar qual o melhor rearranjo das duas sequências em estudo. Nessa determinação são também tidas em conta as zonas de uma determinada sequência desemparelhadas.

Nemática, fase: Fase líquida cristalina, em que as moléculas não têm ordem posicional mas possuem ordem segundo uma determinada orientação.

Pericíclica, reacção: É uma classe de reacções orgânicas em que o estado de transição tem uma geometria cíclica, havendo rearranjo concertado de átomos e electrões, sem que surja na sequência algum composto intermediário. Os principais tipos de reacções pericíclicas são: electrocíclicas, cicloadições, reacções sigmatrópicas e quelotrópicas. A reacção mais conhecida e ilustrativa deste tipo de reacções é a reacção de Diels-Alder:



Polimorfismo: É a capacidade de um material sólido de coexistir em mais de uma forma ou estrutura cristalina.

Pontes de hidrogénio: Classe especial de forças de Van der Waals, nomeadamente do tipo dipolo-dipolo. Este tipo de interações dá-se entre um átomo electronegativo e um átomo de hidrogénio ligado a outro átomo electronegativo (nomeadamente azoto, oxigénio e flúor).

Ponto de ebulição, (*p.e.*): É a temperatura em que a pressão do líquido iguala a pressão atmosférica e o líquido entra em ebulição.

Rede cristalina: Conjunto de pontos que se repetem indefinidamente, cada ponto é um cristal, constituído por um átomo ou uma série de átomos.

Refracção molar, (R_M): É uma medida do volume ocupado por um átomo ou grupo de átomos e está definido matematicamente por: $R_M = \frac{N_A \alpha}{3\epsilon_0}$, em que N_A é o número de Avogadro, α a polarizabilidade e ϵ_0 a permissividade do vácuo. Tem como unidades o $\text{m}^3 \cdot \text{mol}^{-1}$.

Regressão, modelo: Modelo genérico que se baseia na relação entre as variáveis de *input* (independentes) e as variáveis dependentes, estas últimas são variáveis numéricas com uma distribuição aproximadamente contínua.

Regressão Multilinear (MLR): Método que estabelece uma relação linear entre uma variável y e uma série de variáveis x_i com a seguinte forma: $y = a_1x_1 + a_2x_2 + \dots + a_nx_n + a_0$ em que os parâmetros $a_0, a_1, \dots, a_2, a_n$ são coeficientes otimizados pela minimização dos quadrados das diferenças entre o valor de y_{exp} experimental e o valor previsto y_{prev} .

Retropropagação, redes neuronais (BPG-NN): Método de aprendizagem automático, cujas unidades fundamentais, os neurónios, se organizam em diversas camadas, camada de *input*, camada oculta (esta podendo conter diversos níveis) e camada de *output*. Quando é submetido um objecto para obter previsões, passa primeiro por cada um dos neurónios da camada de *input*, cada um desses neurónios está ligado a cada um dos neurónios da camada imediatamente abaixo (neste caso na camada oculta), cada uma dessas ligações é feita através de um determinado peso w_{ji} . Esses pesos serão usados numa combinação linear que depois serve para determinar *output* desse neurónio (neste caso da camada oculta), função de todos os *outputs* da camada acima e respectivos pesos w_{ij} associados às ligações dos neurónios da camada acima com o neurónio em causa. Esse procedimento vai-se repetir à medida que se aproximam os neurónios da camada de *output*, sendo assim obtida previsão (*output*) para determinado objecto. No processo de treino o sentido é inverso, servindo os desvios entre valores previstos (*output*) e valores experimentais na camada de *output* para corrigir os pesos que ligam os neurónios da camada de *output* aos neurónios da camada imediatamente acima; esses pesos servem para calcular desvio para *output* de neurónios dessa camada acima. Repetindo-se esse procedimento para as camadas mais próximas da camada de *input*. Propaga-se assim a correcção dos erros em sentido contrário (retropropagação) ao da obtenção de previsões pelo modelo.

Supervisionado, método: É um tipo de aprendizagem automática que na sua fase de treino, estabelece uma relação entre os parâmetros de *input* e os de *output* (propriedade ou propriedades a determinar).

Tanimoto, Coeficiente: Aplicação específica do coeficiente de Jaccard para medidas de semelhança entre vectores binários. Genericamente o coeficiente de Jaccard é definido por:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Para o caso da comparação entre vectores binários a equação genérica representada acima, assume a seguinte forma como o coeficiente de Tanimoto:

$$Tanimoto(A, B) = \frac{N_{AB}}{N_A + N_B - N_{AB}}$$

Em que N_{AB} é o número de bits que os vectores A e B compartilham com valor idêntico a 1, N_A é o número de bits que o vector A tem igual 1, e N_B o número de bits que o vector B tem igual a 1.

Termotrópico: Referente a um composto que exibe uma transição de fase para uma fase líquida cristalina com mudança de temperatura.

Validação cruzada (c.v.): É a pratica estatística de partir um conjunto de dados em subconjuntos, uma parte desses subconjuntos é utilizado para construir um modelo, a outra parte é utilizada para validar o modelo construído. Na validação de *Holdout* os mesmos subconjuntos permanecem como conjunto de treino e conjunto de teste, Na validação cruzada K x formam-se K subconjuntos, um deles permanece fora para validação, enquanto os outros $K-1$ objectos são usados para construir o modelo, esse procedimento repete-se até todos os K objectos terem sido conjuntos de validação. Na validação *leave one out* o número de subconjuntos é igual ao número total de objectos, ficando de cada vez um objecto de fora para validação enquanto os restantes objectos são usados para construir o modelo.

Van der Waals, forças: Forças atractivas ou repulsivas entre moléculas e incluem interacções dipolo-dipolo, dipolo permanente-dipolo induzido e interacções entre dipolos instantâneos induzidos (forças de London).

Xenobióticos, compostos: São compostos que, apesar de não serem produzidos no organismo, estão aí presentes. Podendo também ser compostos que estão em concentrações muito mais elevadas do que é habitual. Fármacos, como antibióticos, são exemplos de compostos xenobióticos.